THE
PROCEEDINGS
OF
THE PHYSICAL SOCIETY

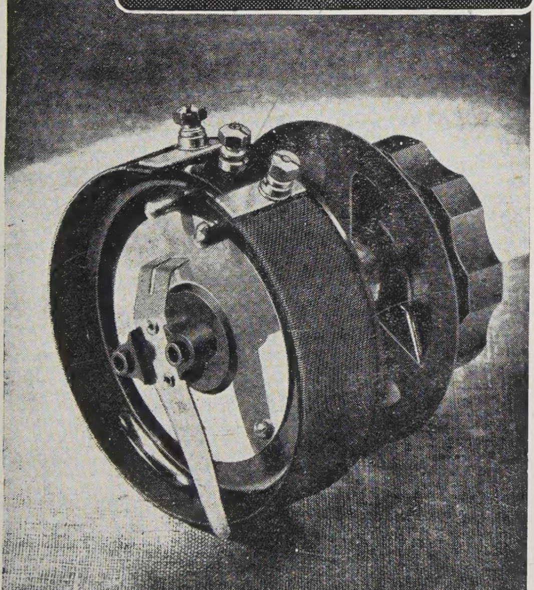
VOL. 55, PART 2 | 1 March 1943 | No. 308

CONTENTS

Price to non-Fellows 8*s*. 4*d*. net; post free 8*s*. 9*d*.
Annual subscription 42*s*. post free, payable in advance

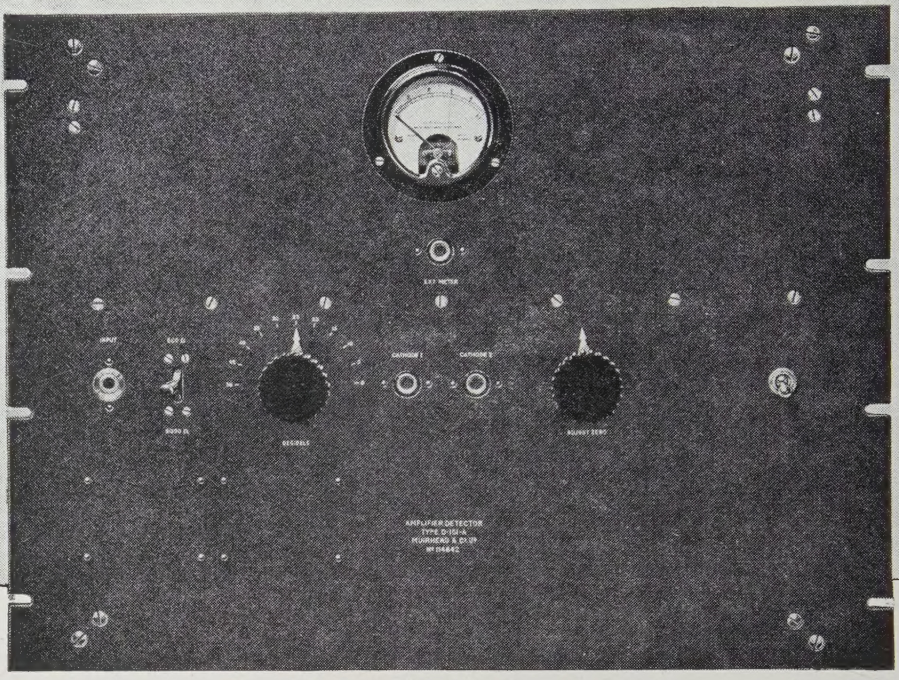
Published by
THE PHYSICAL SOCIETY
1 Lowther Gardens, Exhibition Road, London, S.W.7

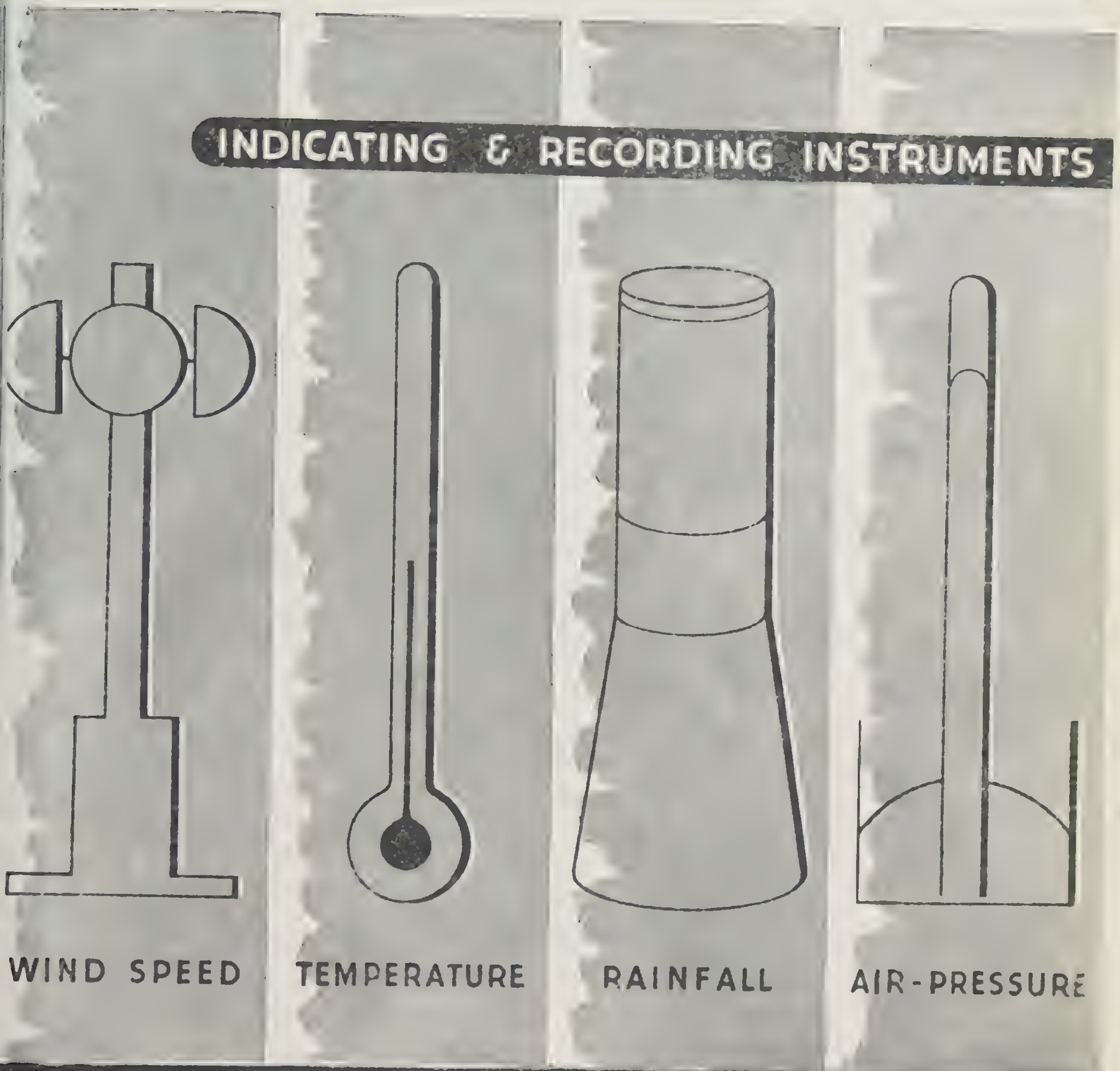
Printed by
TAYLOR AND FRANCIS, LTD.,
Red Lion Court, Fleet Street, London, E.C.4



BERCO Wire-wound Potentiometer


FIRTH - BROWN
INSTO
SEGMENTAL
SAWS
The Firth-Brown Insto Segmental Saw meets the demand for a general purpose saw, capable of dealing rapidly with either light or heavv structural sections or bars. The tooth segments are interchangeable and renewable so that teeth, when worn out, can soon be replaced.
The FIRTH-BROWN HARDOMETER
An instrument combining a high degree of accuracy and reliability, which has been developed and perfected to meet the urgent demand for a reliable hardness tester.
The Hardometer is constructed with the precision of a scientific instrument yet is sufficiently robust to withstand the wear and tear of continuous usage under workshop conditions.
THOS FIRTH & JOHN BROWN LT




INDICATING & RECORDING INSTRUMENTS
WIND SPEED
TEMPERATURE
RAINFALL
AIR-PRESSURE

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 55, PART 2 — 1 March 1943 — No. 308

A PHOTO-ELECTRIC TRICOLORIMETER

Communication H 861 from the Kodak Research Laboratories

By G. F. G. KNIPE and J. B. REID

Paper read to the Colour Group 11 *September* 1942

ABSTRACT. The colorimeter described was designed to measure the colour of nearly-white papers. It makes use of a double monochromator system and three diaphragms which select the correct proportion of light of each wave-length to enable the result to be given in C.I.E. units directly.

§ 1. INTRODUCTION

THE instrument described in this report was designed to measure the colour of nearly-white paper with sufficient accuracy to specify those differences which are visible, but the nature of which cannot be determined by visual colorimeters. As an accuracy of the order of ten times that normally obtained on visual colorimeters is required, a photo-electric method was indicated. In view of the almost universal adoption of the C.I.E. colorimetric system, it was decided that any photo-electric instrument constructed for the above purpose should approximate as closely as possible to this system, which has the advantage that one of the colour co-ordinates also measures the visual reflection factor of the sample. Thus all the important photometric properties, except gloss, are measured by three numbers only. The optical part of this instrument (in a slightly different form) was shown at the Exhibition of the Physical Society in 1939.

The instrument does not measure colour in any physiological sense, but merely calculates the colour from spectro-photometric data in the following way.

Let E_λ be the energy of a standard light source in the band of unit wave-length width at a wave-length λ, and let R_λ be the reflection of a sample for a wave-length λ. Then the tri-stimulus values for the sample illuminated by this light source are

$$X = \int_0^\infty E_\lambda R_\lambda \bar{x}_\lambda \, d\lambda,$$

$$Y = \int_0^\infty E_\lambda R_\lambda \bar{y}_\lambda \, d\lambda,$$

$$Z = \int_0^\infty E_\lambda R_\lambda \bar{z}_\lambda \, d\lambda,$$

where $\bar{x}_\lambda$, $\bar{y}_\lambda$, $\bar{z}_\lambda$, the tri-stimulus values of the spectrum colours, vary with the wave-length and have been given standard values (Commission Internationale, 1931).

If the sample is now illuminated by a light source whose energy (per unit wave-length) at a wave-length λ is E'_λ, so that the reflected energy at λ is $E'_\lambda R_\lambda \delta\lambda$, and this energy passes through a filter whose transmission at a wave-length λ is $T_{x\lambda}$, and falls on a photo-cell, the current from which is proportional to the incident energy, and for unit incident energy of wave-length λ is S_λ, the current from the cell for wave-length λ will be

$$E'_\lambda R_\lambda T_{x\lambda} S_\lambda \,\delta\lambda,$$

and if there is no interference between light of different colours, the total current will be

$$\int_0^\infty E'_\lambda R_\lambda T_{x\lambda} S_\lambda \,d\lambda.$$

Therefore, if we can make $E_\lambda T_{x\lambda} S_\lambda / E_\lambda X_\lambda$ constant for all values of λ, the current from the photo-cell will be proportional to X. Given two other filters $T_{y\lambda}$ and $T_{z\lambda}$, the currents from the photo-cell will be proportional to X, Y and Z.

As photo-cells can be obtained of sufficient linearity, constancy of colour response, and with arithmetic summation of the currents set up by light of different colours, the main difficulty lies in the manufacture of suitable filters. Two methods are available. In the first, optical filters are used, and in the second the light is resolved into a spectrum, the correct amount of light of each wave-length being selected by a metal diaphragm placed in the plane of the spectrum (Newton, 1730). While the first method is more sensitive, i.e. passes more light, as the light does not have to go through a slit, the difficulty of obtaining sufficient accuracy with filters to read C.I.E. units is very great (Féry, 1908). The present instrument, therefore, uses the second method, which was suggested by Strache (1911), and has been used by several other authors (Ives, 1915; Koenig, 1934, 1937; Voogd, 1939; Winch and Machin, 1940). A combination of the two methods may be used when it is wished to reduce the effects of stray light, but the amount of stray light in this instrument is very small, owing to the use of a double monochromator system.

§ 2. DESCRIPTION OF APPARATUS

(a) *Optical system*

The optical part of the apparatus consists of two spectroscopes so arranged that the light incident on the slit S_1 of the first spectroscope is dispersed into a spectrum and is then recombined by the second into white light, which passes through the final slit S_2 on to the specimen being measured (figure 1). The reflected light from the sample is picked up by two photo-cells which receive the light reflected at an angle of approximately 45° from the sample. A double system has been preferred to a single spectroscope with a lens over the spectrum to recombine the light, because it gives greater freedom from stray light. This is shown by the fact that the instrument can be used in daylight with the cover removed from the first spectroscope. The existence of an appreciable amount of stray light would make it impossible to reduce the transmission of the instrument to the small values required in some parts of the spectrum. Diaphragms have been provided around all lenses to reduce stray light as much as possible.

It is desirable that the distribution of light in the spectrum should be fairly even, otherwise sudden jumps in the outline of the diaphragm would be necessary.

Therefore, the entrance slit must be evenly illuminated, for the spectrum is an image of this slit. Moreover, it is desirable that the illumination of the sample should be even, otherwise undue weight is given to certain areas of the sample. Consequently the light source itself must be of uniform brightness. Despite certain disadvantages, therefore, a ribbon-filament lamp R has been used as a light source, and a magnified image is thrown on the entrance slit by the lens L_1. This makes for stability against small changes of position and shape of the filament, as well as for uniformity of illumination of the entrance slit. The lamp takes 108 watts at 6 volts, and the filament is about 2 mm. by 10 mm. It is run from three 8-volt car accumulators which are connected in parallel for running and in series for recharging. A potentiometer and standard cell are used to measure the voltage drop on a standard resistance of about 1/10 ohm, through which the lamp current flows, thus giving a sensitive indication of the constancy of the lamp current. This voltage must be kept constant to within a millivolt, and so a very smooth current control is necessary. This is obtained by using a high

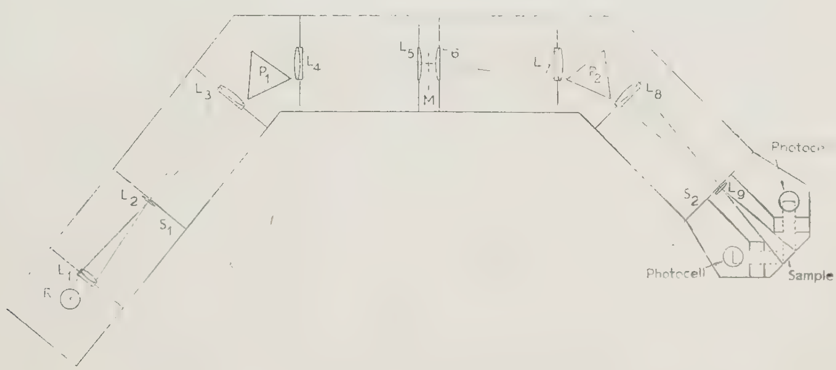


Figure 1. Diagram of optical system of colorimeter.

variable resistance in parallel with a small fixed one, the values being approximately 20 ohms and 0·25 ohm. The high resistance has a hyperbolic winding in order to give equal sensitivity of adjustment at all points on it.

The entrance and exit slits, which are about 0·5 mm. wide and 2·25 cm. long, were made by removing a line of silver from an ordinary mirror and cementing a cover glass on the side from which the silver has been removed. As there is a strong reflection from both sides of these slits, it is necessary to blacken the glass surface up to as near the edge of the slit as possible. This type of slit was adopted because it was thought likely to remain permanently of the same width, could be cleaned simply, and did not provide an entrance for dust.

A field lens L_2 throws an image of the condenser lens L_1 on to the collimator lens L_3. The parallel light from this lens is dispersed by the prism P_1 and brought to a focus at M, the position of the diaphragm, by the lens L_4, P_1 being adjusted so that the spectrum is level and the lines are parallel to the slit. The field lenses L_5 and L_6 throw an image of L_4 on L_7, so that all the light from the spectrum passes through L_7, which collimates the light before it passes through the prism P_2, and is brought to a focus on the slit S_2, through which it passes to the sample. A standard magnesium-oxide surface can be swung up in front

of the sample by turning a handle. The transparent samples are placed over the first collimating lens L_3. The beam covers a fairly large area here, and so an overall measurement of the colour is obtained. In addition, if the filter is slightly tilted, any light reflected from the surfaces will not pass through slit S_2 on to the magnesium-oxide surface which is used for transmission measurements, and the position of the filter is then not critical.

The diaphragms are mounted in a metal frame moving in a slide which is designed on kinematic principles and arranged perpendicular to the axes of lenses L_5, L_6. The limits of travel of the slide are fixed by two adjustable stops which position the slide for measurements on the X and Z diaphragms. A spring plunger engages in a slot in the frame for positioning during measurements on the Y diaphragm.

The spectrum is somewhat larger than the diaphragm, so that movements of the diaphragm in a vertical plane will not cause errors. However, the diaphragm must not move in a horizontal direction, or light of a given wavelength will fall on the wrong part of it, and location in this direction is critical.

(*b*) *Electrical system*

The light reflected from the sample is received by two high-vacuum thin-film potassium-on-silver photo-cells, which are mounted together with the electrometer valve and the grid resistance in the housing of the instrument. The rest of the electrical circuit is mounted in the table and is connected to the instrument by a six-way screened cable and connector.

The amplifier is a form of the DuBridge and Brown circuit (1933), modified to use a standard American valve, the 1A6 (figure 2) (Coven, 1938). It is essentially a bridge circuit with two grids in the valve acting as anodes and forming two of the arms. Since both these grids receive their electrons from the same filament, they are affected to the same extent by fluctuations in filament emission, and the galvanometer is undisturbed.

Another important feature of the circuit is that for a certain value of the filament current the galvanometer reading passes through a maximum, and hence remains stationary for small fluctuations in the filament current. The resistance r_6 ensures that the galvanometer is shunted by an approximately constant resistance and is not underdamped when r_4 is at a position of low sensitivity (Reid, 1942). With the light available, a maximum potential of about 1·0 volt is obtained across an input resistance of 500 megohms, with which the amplifier is quite stable. During calibration, however, a resistance of 5000 megohms is necessary, and with this resistance there was some instability, and frequent checks were necessary to eliminate errors due to this.

§ 3. CALIBRATION

The shapes of the diaphragms were determined empirically by the following process, which avoids all difficulties due to lack of knowledge of the energy distribution of the incident light, light loss in the apparatus, and the colour sensitivity of the photo-electric cells.

A slit having the mean curvature of the spectrum lines was inserted in the spectrum. This slit could be moved to any desired position of the spectrum

by means of a micrometer, and its length could be adjusted symmetrically about the centre line of the spectrum by a second micrometer. Then by trial the length of slit was found which gave a photo-current determined by the equation

$$C_{z\lambda} = K\left(E_{\mathrm{B}}\bar{z}\,\frac{d\lambda}{dm}\right)_{\lambda}\Delta m,$$

where $C_{z\lambda}$ is the required photo-current for the Z diaphragm when the slit is set at wave-length λ, K is a constant chosen to give the maximum possible diaphragm opening at the wave-length where the opening is greatest, E_{B} is the energy distribution of the C.I.E. illuminant B, $\frac{d\lambda}{dm}$ the dispersion of the spectrum and Δm the width of the slit. The height so determined is that for the Z diaphragm at

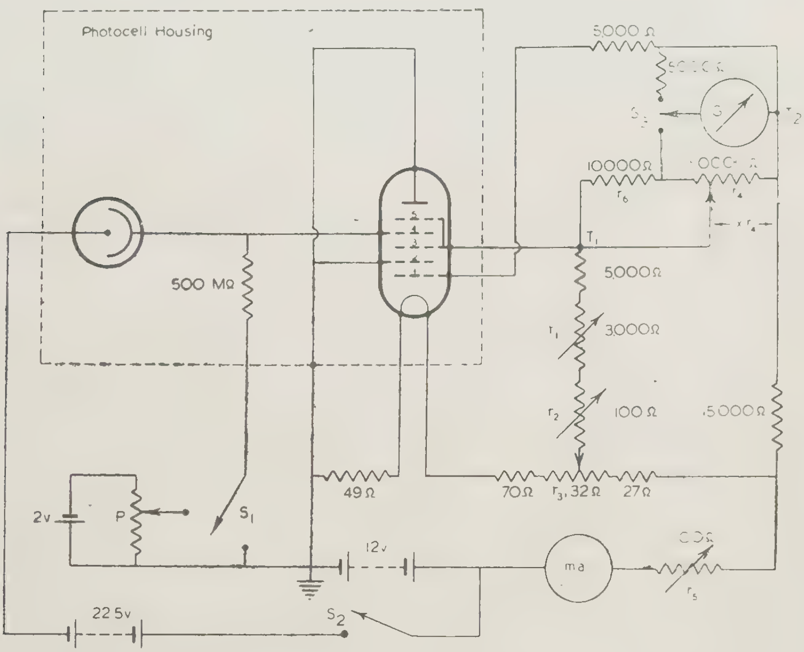


Figure 2. Electrical circuit of amplifier.

wave-length λ. Similar equations, with the same K, apply to the X and Y diaphragms.

The wave-length calibration was carried out using a mercury-cadmium discharge lamp as a line-spectrum source and passing the calibrating slit, which runs on a slide just in front of the diaphragm slide, through the spectrum. The micrometer reading was taken when a line just appeared on one side of the slit and when it was just about to disappear on the other, the mean of these two readings being taken as the micrometer reading for the wave-length of that line. This was done for eight lines, and a Hartmann formula calculated which fitted the readings within $\pm 0{\cdot}6\,\mathrm{m}\mu$. The micrometer readings and dispersions corresponding to wave-lengths from $400\,\mathrm{m}\mu$ to $720\,\mathrm{m}\mu$ at intervals of $5\,\mathrm{m}\mu$ were calculated from the formula, and the values of $E_{\mathrm{B}}\bar{x}\,\frac{d\lambda}{dm}$, $E_{\mathrm{B}}\bar{y}\,\frac{d\lambda}{dm}$ and $E_{\mathrm{B}}\bar{z}\,\frac{d\lambda}{dm}$

were worked out for each of these wave-lengths. These numbers represent the photo-current which should be produced at these points in the spectrum, and when plotted they give the curves shown in unbroken lines, figure 3.

In determining these values the X, Y, and Z diaphragms were made correct relatively to each other by determining their openings at a small number of wave-lengths, and the shape of each diaphragm was then determined by interpolation between these values. This was desirable because of the instability mentioned above, and excellent agreement was found between repeated estimates of the shape of the diaphragms when this method was adopted. The diaphragms were then cut from sheet brass on a milling machine, using

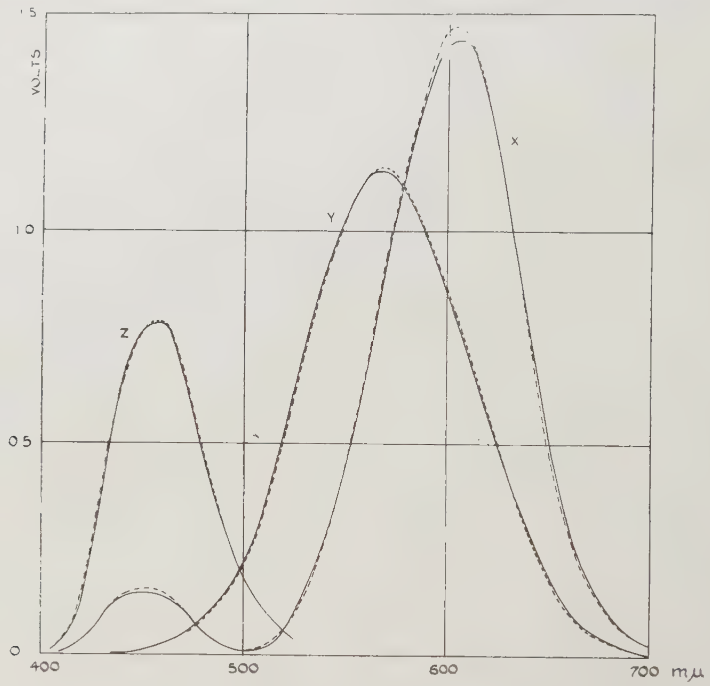


Figure 3. Calculated and measured diaphragm transmissions.

the feed micrometers to set the cutter accurately. Each opening is 0·85 in. long and the maximum width is about 0·5 in.

The transmission of the diaphragms was checked by placing the calibration slit, the jaws being wide open, at each of the positions in turn at which the values of the transmission had been calculated, and recording the resulting photo-currents. The transmission of the X diaphragm was low in the region from $\lambda = 400$ mμ to $\lambda = 500$ mμ, and this was corrected by filing, frequent checks being made to avoid taking off too much. The final results are shown by the curves with a broken line. The ordinates of the curves have been multiplied by a factor to make the total transmission equal to the theoretical total transmission, so that the curves may be compared more easily.

The accuracy of the instrument might be increased and the position of the diaphragms made less critical if the two photo-cells had differing spectral

sensitivities, chosen to increase the width of the diaphragms at the points where they are, at present, rather narrow.

§ 4. PERFORMANCE

To test the general accuracy of the instrument, seven Wratten filters were chosen, which, from the shape of their transmission curves, would provide the most critical test that filters could provide. Transmission samples were chosen in preference to opaque samples because they were less liable to get soiled or

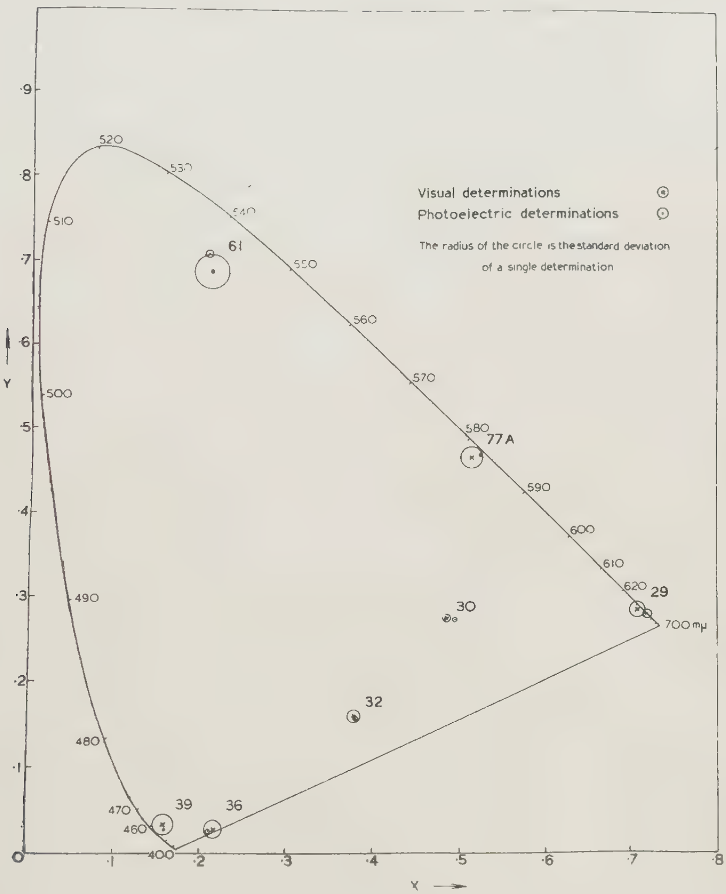


Figure 4. C.I.E. colour diagram showing visual and photo-electric measures of certain filters.

change in colour over a period. These filters were measured on the photo-electric instrument and also by six observers on the Donaldson colorimeter, which is a visual instrument.

The results are shown in the colour chart of figure 4. The agreement between the photo-electric instrument and the Donaldson colorimeter is satisfactory, indicating a satisfactory degree of accuracy in the masks. The precision of the photo-electric instrument is about five times that of the Donaldson on these saturated colours.

When used on paper, the accuracy is considerably higher, as a check can be made on the standard white surface, which is close to the papers on the colour

chart. As the reflection factor does not vary rapidly through the spectrum, any inaccuracy in the diaphragm causes correspondingly less error in the results. The main inaccuracy has been found to be caused by slight differences in the setting of the diaphragms, and this has been avoided by reading the transmission through the X diaphragm first on the white surface, then on paper, and then again on the white surface. If the first and third readings are less than 2 millivolts apart, the measurement on the X diaphragm is taken as satisfactory and readings are made on the Y and Z diaphragms. In this way the diaphragm setting is not altered between readings, and high precision is obtained. This is shown in the table below, where the first two columns give the colour coefficients and their standard deviations, and the third gives the reflection factor and its standard deviation. Three sets of measurements were made on three different sheets of each paper, so that the deviations given below include any variation in the paper.

Table of measurements on typical photographic papers (white and cream)

Papers	x	Δx	y	Δy	Y	ΔY
A (white)	0·3347	0·0002	0·3354	0·0001	0·825	0·000
B (white)	0·3352	0·0002	0·3360	0·0001	0·797	0·002
C (white)	0·3359	0·0001	0·3364	0·0003	0·876	0·001
D (white)	0·3367	0·0001	0·3363	0·0001	0·861	0·001
E (white)	0·3369	0·0001	0·3375	0·0001	0·864	0·002
F (white)	0·3465	0·0002	0·3479	0·0002	0·864	0·001
G (cream)	0·3523	0·0000	0·3538	0·0002	0·843	0·001
H (cream)	0·3655	0·0003	0·3622	0·0002	0·796	0·006

These deviations lie well within the limits of 0·0005 for colour and 0·005 for reflection factor which were mentioned at a recent meeting of the Colour Group of the Physical Society as being necessary for industrial measurements (Wright, 1941). Small differences can, therefore, be measured with precision, and the absolute accuracy is also very satisfactory.* Owing to the difficulty of making measurements on paper with a visual colorimeter, these samples have not been measured visually. Recent work (MacAdam, 1942) has shown that the standard deviation of visual measurements on near-whites under good conditions is about 0·0020.

The early experiments which led to the construction of this instrument were carried out by Dr. F. H. G. Pitt, who demonstrated the feasibility of constructing the instrument and the accuracy which was possible.

REFERENCES

Brown, H. and DuBridge, L. A., 1933. *Rev. Sci. Instrum.* **4,** 532.
Commission Internationale de l'Eclairage, 1931. *Procès Verbaux.*
Coven, A. W., 1938. *Rev. Sci. Instrum.* **9,** 229.
Féry, C., 1908. *J. Phys. Chem.* **7,** 632.
Ives, H. E., 1915. *Phys. Rev.* **6,** 337.
Koenig, H., 1934. *Helv. Phys. Acta,* **7,** 433.
Koenig, H., 1937. *Helv. Phys. Acta,* **10,** 165.
MacAdam, D. L., 1942. *J. Opt. Soc. Amer.* **32,** 247.
Newton, I., 1730. *Opticks,* First Book, Part 2 (Reprint, Bell, 1931).
Physical Society, 1939. *Catalogue of the 29th Annual Exhibition,* p. 220.

* See note by H. M. Ross on page 90.

REID, J. B., 1942. *J. Sci. Instrum.* **19**, 124.
STRACHE, H., 1911. *Proc. Amer. Gas Inst.* **2**, 401.
VOOGD, J., 1939. *Philips' Tech. Rev.* **4**, 260.
WINCH, C. T. and MACHIN, C. F., 1940. *Trans. Illum. Engng. Soc., Lond.*, **5**, 93.
WRIGHT, W. D., 1941. *Nature, Lond.*, **148**, 506.

DISCUSSION

Mr. R. G. HORNER. Was the main purpose in having two photo-cells to get double the sensitivity, or to average the spectral sensitivities of the cells, or to average the light reflected at the angle of 45° ?

Mr. J. H. PRENTICE. In connection with the reception system chosen, I notice this is the same type as that used in the colorimeter described by Winch and Machin (G.E.C.). I should like to know if this was chosen for any specific reason and whether any experiments with other types of photo-electric cell were conducted, with a view to establishing the suitability of this particular type.

Mr. F. HAEGELE. I note in this apparatus certain similarities to an instrument described by P. M. Van Alphen in *Philips' Technical Review* a few years ago. A double monochromatic system was used in conjunction with a photo-cell and lamps in an integrating sphere. Instead of diaphragms for C.I.E. coefficients, 8 blocks were selected for measuring the colour rendering of light sources. This worker produced the diaphragm by a photographic method. Have Messrs. Kodak tried this, and if so, what were its disadvantages ? Since the amount of light is small, was a large aperture optical system used ?

In the colour chart of filter measurements I see No. 77A for Hg $\lambda = 5461$ A. Why is its colour point about 5850 A. ?

What protection is there against ageing of the lamp affecting the colour temperature ?

Dr. T. VICKERSTAFF. The instrument was designed so that the diaphragms converted the ribbon-filament lamp illumination to C.I.E. results in illuminant B. Would it not be possible to use diaphragms giving results in A so that the C.I.E. solutions could be inserted in front of the lamp to give results in B and C, thus rendering the instrument more versatile ?

Mr. R. W. KERSEY. Can the authors give information as to (i) stability of spectral characteristic of potassium vacuum cell employed ; (ii) voltage stabilizing device for lamp supply ; (iii) degree of similarity between published data on gelatine filters and results of measurement ? Has any work been done on other methods employing colour-corrected photo-cells, and have any special filters been developed ? Was the second method adopted because the filter method seemed so hopeless ?

A special ring-shaped cell for picture transmission is made in Germany (Siemens-Karolus) which picks up an annular cone of light. This might be of some value in the colorimeter to replace the two photo-cells.

Mr. P. S. H. HENRY. Would it be possible to get over the difficulty of making the diaphragms sufficiently accurate at the points where they were narrow by using them in conjunction with suitable filters ? These filters, which would not have to possess accurately specified characteristics, would absorb in those regions of the spectrum where the present diaphragms have to be made narrow, so that the new diaphragms would not have to be narrow at any point.

Dr. W. D. WRIGHT. It is very interesting to find that the colorimeter described in this paper combines both a high sensitivity and a reasonably good absolute accuracy. As the authors show, the absolute accuracy is also high when measuring white surfaces, if what is in effect a substitution method is used, in which the instrument merely measures the difference in colour between the surface under test and a standard magnesium-oxide surface. Do the authors consider that the substitution method could be extended to cover the whole of the colour field to the same order of accuracy ? I should also like to ask the authors whether an increase in scale of the apparatus would be feasible, and would,

by increasing the dimensions of the templates, result in a higher accuracy ; and whether there is any likelihood of the apparatus being put on the market, as it would no doubt prove a valuable tool in a number of industries.

Mr. H. M. Ross (communicated). Since this paper was written and presented, it has been possible to carry out tests to check the absolute accuracy of this tricolorimeter. Three white and two cream-coloured photographic papers were measured on the Hardy Recording Spectrophotometer by the Eastman Kodak Research Laboratories, at Rochester, U.S.A., and their colours computed for the C.I.E. illuminant B. These same samples were also measured on the tricolorimeter under as nearly as possible the same conditions, the results being given in the table. The first row of figures for each paper gives the Hardy Spectrophotometer results, the second the mean of our determinations (four in number, taken over a period of several days), the third the mean of the differences between each of our determinations and the mean values given in the second line.

Paper	Test	Tristimulus values			Trichromatic coefficients	
		X	*Y*	*Z*	*x*	*y*
Blue-white	Rochester	0·8415	0·8466	0·7102	0·3509	0·3530
	Harrow	0·8187	0·8235	0·6913	0·3510	0·3528
	Mean of Harrow diffs.	0·0005	0·0002	0·0002	0·00008	0·00003
Natural	Rochester	0·9159	0·9238	0·7384	0·3553	0·3583
	Harrow	0·8867	0·8955	0·7184	0·3547	0·3581
	Mean of Harrow diffs.	0·0004	0·0007	0·0002	0·00015	0·00008
Cream-white	Rochester	0·8738	0·8774	0·6749	0·3602	0·3617
	Harrow	0·8498	0·8545	0·6567	0·3597	0·3618
	Mean of Harrow diffs.	0·0003	0·0002	0·0001	0·0002	0·00008
Ivory	Rochest	0·8518	0·8472	0·5449	0·3796	0·3776
	Harrow	0·8223	0·8204	0·5278	0·3789	0·3779
	Mean of Harrow diffs.	0·0003	0·0005	0·0002	0·00008	0·00008
Buff	Rochester	0·8593	0·8616	0·5580	0·3771	0·3781
	Harrow	0·8319	0·8368	0·5398	0·3767	0·3790
	Mean of Harrow diffs.	0·0002	0·0004	0·0001	0·00005	0·00008

A test was also made to find the effect of scanning different areas of one sample. The blue-white, which has a smooth surface, was chosen and moved not more than half an inch from the normal scanning position. Four measurements were taken in succession, the mean of the differences being :

X : 0·00035 Y : 0·00043 Z : 0·00035 x : 0·00015 y : 0·00003.

Improvements have also been made, since the paper was presented, to the amplifier and lamp-controlling circuits of the tricolorimeter. A test was therefore carried out to redetermine the precision, or repetition accuracy, of the instrument. The blue-white paper was measured eleven times, over a period of several days, care being taken not to move the sample during this time. The means of the differences of the eleven determinations are :—

X : 0·00015 Y : 0·00012 Z : 0·00012 x : 0·00003 y : 0·00002.

It would appear, therefore, that for the measurement of the colour of near-white papers the precision of the tricolorimeter is fully satisfactory, and that the absolute accuracy of colour determination, relative to the spectrophotometric measurements, is within acceptable

tolerances, particularly when taking into account the possibility of scanning slightly different areas of the samples. It will be noticed, however, that the present values for X, Y and Z are consistently lower than the spectrophotometric figures, but this is almost certainly due to the tricolorimeter measuring only the light reflected at approximately 45° from the sample, whereas the spectrophotometer uses an integrating sphere to measure all the reflected light.

AUTHORS' reply. In reply to Mr. R. G. Horner, the main purpose in having two photo-cells was to increase the sensitivity, and they also offer an opportunity of controlling the spectral sensitivity to some extent.

In reply to Mr. J. H. Prentice, the first model of the colorimeter made use of an integrating sphere, but this was altered as it was found that false colour readings would be obtained due to multiple reflections in the sphere. The photo-cells used were recommended by the G.E.C. and have proved to be quite satisfactory.

As regards the point raised by Mr. F. Haegele, a photographic method was tried, but light scatter in the emulsion caused trouble where the diaphragms were narrow. In addition, the gelatine in the clear parts of the image is seldom quite colourless, and the silver might not be sufficiently opaque to the large amount of infra-red radiation present. Also, if any adjustment to the diaphragm is found necessary, a completely new one has to be made. The aperture of the lenses used was about F/5, which was the largest available at the time. The Wratten 77A filter transmits a wide band of red in addition to the narrow band at 5461 A., and this shifts its colour towards the red. As there is little red in the mercury spectrum, this wide band is immaterial when isolating the Hg 5461 A. line. The lamp used has some grinding material inside which can be used to remove the film of tungsten which collects on the inner surface of the glass. The check on the coefficients for the standard white will show if the colour temperature of the lamp has altered, and the current may have to be altered to compensate for this.

We agree that Dr. T. Vickerstaff's suggestion that diaphragms for illuminant A might be used with the standard C.I.E. solutions seems to be a good one if only one set of diaphragms is to be made.

In reply to Mr. R. W. Kersey, no serious changes in the spectral sensitivity of the cells used have been found over a period of three years. No stabilizing device has been used with the lamp, owing to the large current to be controlled. A stabilizing transformer might be used with A.C., but the measurement of the lamp current would then be more difficult. We believe that the tolerances allowed on gelatine filters are different for different filters, but we have no knowledge of what the tolerances are. Owing to difficulties encountered by the other workers, it was decided to omit filters altogether.

It is probable that filters could be found to improve the shape of the diaphragms, as suggested by Mr. P. S. H. Henry, although no experiments were made in this direction.

In reply to Dr. W. D. Wright, it seems that the substitution method might be extended to cover the whole colour field with considerable accuracy, although probably not with the accuracy obtainable on near-whites. The accuracy and convenience of calibration and operation could be improved by increasing the scale of the apparatus. There is no likelihood of the apparatus being put on the market.

REFRACTIVE INDICES OF GASES AT HIGH RADIO FREQUENCIES

By F. J. KERR, M.Sc., A.Inst.P.,
Melbourne

Communicated by Prof. T. H. Laby, F.R.S. ; *received* 10 *October* 1942

§ 1. INTRODUCTION

In recent years, ultraviolet waves, or wireless waves less than 10 metres in length, have become increasingly important in many applications in which reliable short-distance communication is required. Propagation of these waves round the earth's curved surface takes place by refraction in the atmosphere, by reflection from discontinuities of refractive index in the troposphere, and by diffraction at the earth's surface. An adequate study of the propagation, therefore, requires a knowledge of the refractive indices of gases at very high frequencies. It was thought desirable to test the assumption, made in all previous theoretical work, that the values of these indices are the same as their values at lower frequencies. When this work was begun, no figures were available for the refractive index or dielectric constant of any gas at a frequency higher than about 4 mc./sec. Since then, however, a result for the dielectric constant of water vapour at 42 mc./sec. has been published by Tregigda (1940).

§ 2. METHOD AND THEORY

The principal methods available for the measurement of refractive indices at radio frequencies are the heterodyne and standing-wave methods. The former, being more suitable at low frequencies, has been used in previous work with gases. It was also used by Tregigda in the determination mentioned above. The latter method, which has been used in the present work, only becomes practicable at high frequencies, where it has been frequently used in work on other materials, but not previously with gases, except for some work on ionized gases, whose refractive indices are much larger. The difficulty in applying the method to gases lies in the smallness of their refractive indices, requiring very precise measurements for their determination.

If a length of short-circuited transmission line is coupled to an oscillator, standing waves will be set up along the line, when its electrical length is equal to an integral number of half-wavelengths of the exciting frequency.

The refractive index of a gas, μ, is equal to the ratio of the velocity in a vacuum, $\mathbf{c}$, to the velocity in the gas, v:

$$\mu = \mathbf{c}/v.$$

For a given frequency, this is equal to the ratio of the wavelength in a vacuum, λ_v, to the wavelength in the gas, λ_g,

$$\mu = \lambda_v/\lambda_g.$$

Thus the refractive index is obtainable from a comparison of the length of standing waves set up in the resonant line in a vacuum with their length in the gas.

In this experiment, the two conductors of the line took the form of a pair of coaxial tubes, the outer one of which was earthed, thus ensuring that the system was completely shielded from external disturbances. A coaxial system lends itself more readily to the construction of a gas-tight enclosure than a parallel-wire system, as the gas under consideration can be contained between the two tubes. Moreover, using a coaxial system, the field is entirely across the gas.

The input end of this coaxial system was coupled to a stable ultra-high-frequency generator, and its effective length altered by means of a "shorting" plunger, driven along the line by rotation of the inner tube of the coaxial pair. When the coaxial line came into resonance with the oscillator, standing waves could be set up, the length of which could be measured, first with the tube evacuated, and then with the tube filled with a gas, the ratio of the two wavelengths giving the refractive index of the gas.

The most commonly used methods of detecting standing waves on a transmission line, i.e., those of studying the reaction of the line on the oscillator to which it is coupled, or of measuring the current in the short-circuiting bridge by means of a thermo-ammeter, are unsuitable here, the former because it is desirable for accuracy to avoid too tight a coupling between the oscillator and line, the latter because of its awkwardness in a coaxial system. In this work, the standing waves were investigated by placing a vacuum-tube voltmeter at a point on the line near its input end, the changes in voltage registered by it being studied as a reflecting bridge was moved along the wires. The voltmeter could not be placed right at the end of the line, the input end being a voltage node, but, at a point a small distance from the end, the voltage varies in direct proportion to the current in the reflecting bridge, so that by this method resonance curves are obtained of exactly similar form to those found in the conventional current-in-bridge method.

The presence of the valve voltmeter across the line has the effect of placing a capacity across the system near the input end, thus bringing the positions of the plunger which give current maxima closer to this end of the line, as shown in figure 1. (With an A373 diode voltmeter, in the coaxial system used, the distance to the first current maximum was shortened by about 25 %.) It is shown by Hund (1924) that placing a small capacity across the line close to its input end brings the first current maximum closer to the input end, without affecting the distance between the first and second maxima, which is equal to half a wavelength (see figure 1).

In most of the work described here, a single maximum has been used, since in this way longer wavelengths can be accommodated in a tube of given size, and the measurements are facilitated when only a single maximum has to be studied. By extending Hund's theory, a method was worked out by which the actual half-wavelength could be calculated from the distance between the input end and the first maximum, when this distance is less than a half-wavelength owing to the shortening effect of the voltmeter capacity.

§ 3. APPARATUS

(a) *Coaxial Lecher system.* The design of the coaxial system is shown in figure 2. The line itself comprised two brass tubes, 100 in. in length, of external diameters $4\frac{1}{2}$ in. and $1\frac{1}{2}$ in. (These sizes gave a figure for the ratio of the inner

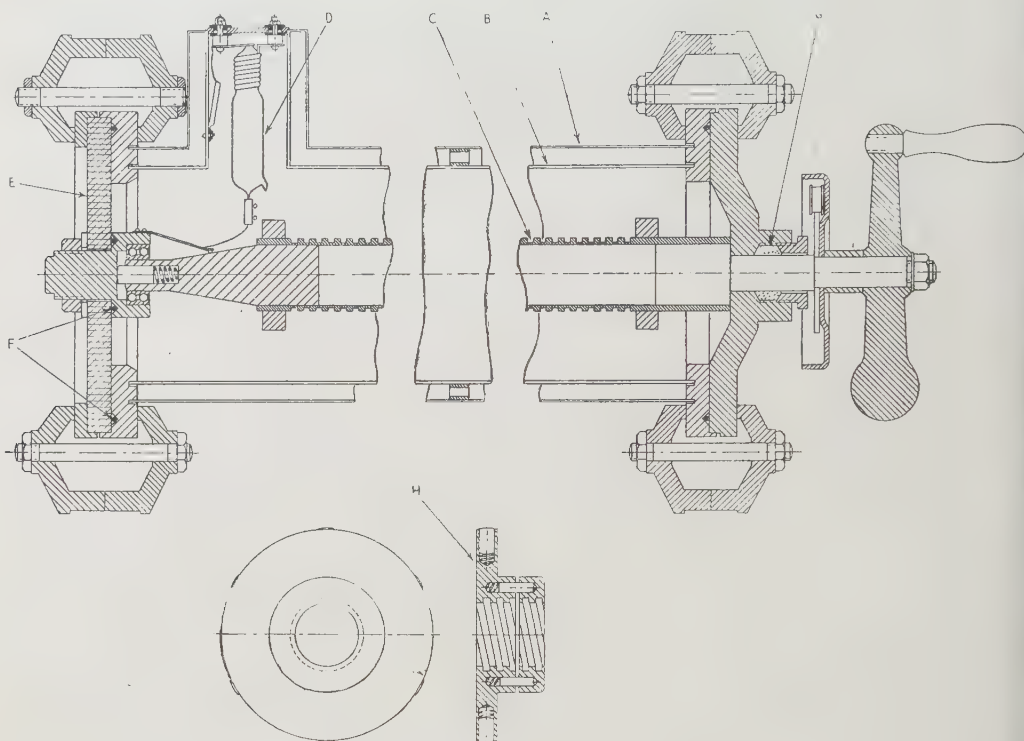


Figure 2.

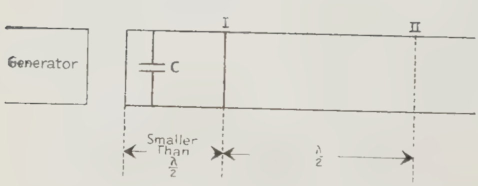


Figure 1. Lecher system, loaded by capacity, showing first two current maxima.

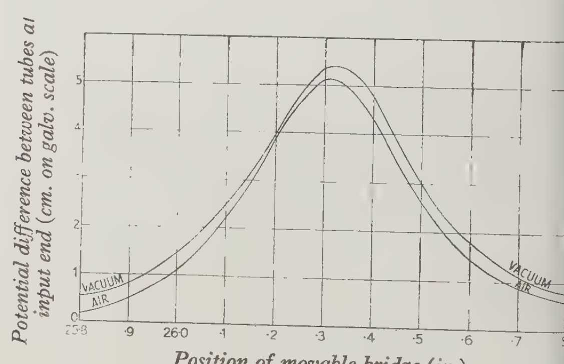


Figure 3.

diameter of the outer conductor to the outer diameter of the inner conductor, which was fairly close to the optimum value, i.e. that which gives the maximum value of Q—defined by the ratio of reactance to resistance—and thus the sharpest resonance curves.)

The two tubes were short-circuited at one point by a plunger, which could be driven along by rotation of the screw-threaded inner tube, thus altering the effective length of the coaxial line.

The ends of the apparatus had to be designed to render the system gas-tight, while carrying the bearings in which the rotatable inner tube turned, and also had to be readily removable for overhauling or adjusting the apparatus. The details of the design can be seen in figure 2.

The input end of the apparatus was closed by an annular piece of $\frac{1}{2}$-in. plate glass, which carried the bearing for the rotatable tube, and also served to insulate the two tubes from each other. The vacuum seals consisted of rings of rubber cord, resting in V-shaped grooves and compressed by the clamps, which held the closing pieces on to the end of the apparatus.

The far end of the coaxial system was closed by a cast bronze domed end-piece, held on by clamps, and also sealed by a compressed rubber ring. The inner tube passed through a gland in this end-piece into a shaft which carried a handle and position indicator. The pitch of the screw was $\frac{1}{2}$ in., and the position of the plunger could be read to 1/1000 in. by means of a revolution counter and a dial indicator. Tests showed that the screw was accurate to nearly 1/1000 in., and that positions of the plunger were reproducible to that amount, the plunger being constructed so that its front face was maintained accurately perpendicular to the axis of the tubes. Stops were arranged, to prevent the plunger from rotating as the threaded tube was turned, so that the plunger would be driven along the tube without rotation. Backlash was minimized in the design, but its effect was eliminated by making all measurements using rotation in the same direction.

A third tube, made of sheet brass, was placed outside the other two and coaxial with them, so that the temperature of the apparatus could be maintained at 100° C. by passing steam through the jacket thus provided, or at a lower temperature by passing water or some vapour through the jacket. The apparatus was lagged by surrounding this tube with several layers of asbestos paper and felt.

(b) *Valve voltmeter.* The valve voltmeter, used for measuring the potential difference between the two conductors near the input end, was located in a short tube at right angles to the main tube, as shown in figure 2. The diode peak voltmeter was chosen as the most suitable at very high frequencies, owing to its high input impedance. This method of voltage measurement has been extensively studied by Megaw (1936), who has published detailed information on the calculation of the errors to which it is subject at high frequencies.

The valve used was the high-frequency diode, type A373. The anode was connected by a flexible wire to a contact wiping on the rotatable inner tube, and the voltmeter condenser was connected between one side of the filament and the outer tube. Care had to be taken to ensure that the terminals for connecting to the battery and galvanometer were gas-tight, while possessing a high resistance to earth, i.e., the outer tube.

(c) *U.H.F. generator.* The refractive indices of gases are so small that some precision is required for their determination. In the apparatus discussed above, the standing waves in the concentric system could be determined to an

accuracy of nearly 1 in 10^6, but to realize this accuracy in refractive-index measurements, it was necessary to use a generator capable of producing a frequency stable to this extent, and to minimize the effects of temperature variations and external disturbances. A line-controlled oscillator having been found insufficiently stable, a crystal-controlled oscillator was used, and for even greater stability the temperature of the crystal was stabilized by placing it in a radiation thermostat, which was a modified form of that described by Laby and Hopper, of this Laboratory (Laby and Hopper, 1939).

Circuits of conventional type were used throughout the generator, in which the frequency was doubled in successive stages from 7·3 mc./sec., the frequency of the crystal, up to 58·3 mc./sec. To eliminate the effect of external disturbances, the various stages were operated inside shielding boxes.

The generator was link-coupled to the coaxial system with a piece of small diameter flexible coaxial cable, several feet in length, which terminated at each end in a single-turn coil, inductively coupled to similar coils in the generator and at the end of the coaxial system. The outside conductor of this coaxial cable was earthed, while the coupling coils were also surrounded by an earthed shield. The whole apparatus was thus completely shielded, a condition which was found to be very necessary for accurate work.

§ 4. RESULTS

(a) *Resonance curves.* Portions of two of the resonance curves obtained are shown in figure 3, in which the potential difference between the tubes at the input end, as measured by the valve voltmeter, has been plotted against the position of the movable bridge. The sharpness of the resonance peaks can be seen from the graph, which is on an expanded scale. By studying the shape of a peak mathematically, the position of a maximum can be obtained to better than 1/1000 in. With the frequency of the oscillator stabilized, the apparatus shielded and the resonance curves so sharp, the limit of accuracy is, in fact, imposed by the irregularities of the screw.

(b) *Dry air.* To check that frequency and temperature drift had been eliminated, each set of determinations was carried out so as to alternate between vacuum and air as rapidly as possible. The procedure was as follows:—After having the apparatus running for a few hours to ensure constancy of temperatures, the tube was pumped out and observations sufficient for drawing a vacuum resonance peak taken. Then air was let into the tube and a run taken with air, after which the tube was again pumped out and a second vacuum run was done, and so on. On each occasion three vacuum and three air runs were taken, thus giving three refractive-index determinations each time. Each run occupied about 10 to 15 minutes, which was the time required to take enough readings to enable an accurate resonance curve to be drawn, while about the same time elapsed between successive runs. The air was dried by passing it through calcium chloride and phosphorus pentoxide.

Three sets of three determinations were made, yielding nine results in all. Table 1 shows one set, taken at a temperature of 100° C. and a frequency of 58·3 mc./sec., which gave three figures for the refractive index of 76 cm. Hg of

dry air. Only one resonance peak was used, allowance being made for the shortening of the distance between the input end of the coaxial system and the first maximum, due to the voltmeter capacity.

Table 1

Run	Mean time (min.)	Pressure (cm. Hg)	Position of max. (in.) vacuum	Position of max. (in.) dry air	Shift (vacuum mean − dry air)	μ (for 76 cm. Hg dry air)
Dry air (i)	17·41	75·9		67·3823	0·0208	$1{\cdot}00021_7$
Vacuum (i)	18·10	0·18	67·4039			
Dry air (ii)	18·43	75·9		67·3779	0·0252	$1{\cdot}00026_2$
Vacuum (ii)	19·8	0·16	67·4021			
Dry air (iii)	19·45	76·0		67·3777	0·0254	$1{\cdot}00026_4$
Vacuum (iii)	20·12	0·17	67·4033			
			(Mean = 67·4031)			

The other two sets of determinations yielded values for the refractive index:

$1{\cdot}00024_5$	$1{\cdot}00022_0$
$1{\cdot}00022_9$	$1{\cdot}00021_7$
$1{\cdot}00023_1$	$1{\cdot}00027_0$

The mean of these nine results, with the probable error calculated on external consistency, is:

Refractive index of dry air at 58·3 mc./sec., 100° c., 76 cm.Hg

$$\mu = 1{\cdot}00023_9 \pm 0{\cdot}00001_4.$$

Assuming the Clausius-Mosotti relation, this is equivalent to a value, for the refractive index of dry air at 58·3 mc./sec., 0° c. and 76 cm. Hg, of

$$\mu = 1{\cdot}00032_7 \pm 0{\cdot}00001_9.$$

(c) *Water vapour.* Water vapour was introduced into the coaxial-tube apparatus, by connecting the latter to a container of water maintained at constant temperature, by a modified form of the radiation thermostat described by Laby and Hopper (1939). With the space over the water evacuated before the water container was heated, this space was subsequently occupied by water vapour alone. Thus the gas inside the concentric system consisted of water vapour at a pressure equal to the saturation vapour pressure, corresponding to the temperature of the water container, as measured by a mercury thermometer in the vapour.

In the early determinations, difficulty was experienced in obtaining equilibrium between the coaxial system and the water tank, and in completely pumping out the tube in readiness for taking a vacuum run. All but two of the results obtained had to be neglected, for one or both of these reasons.

These two gave results for the refractive index of

$$\mu = 1{\cdot}0030_9 \text{ and } 1{\cdot}0029_4,$$

with a mean value for the refractive index of water vapour at 58·3 mc./sec., 100° c. and 76 cm. Hg of

$$\mu = 1{\cdot}0030_1 \pm 0{\cdot}00007.$$

This result shows that the refractive index of water vapour is considerably greater than that of dry air, so that the water vapour in the atmosphere will have a large effect on the refraction of ultra-short waves. The figure agrees well with the value of 1·0060 obtained by Tregigda (1940) for the dielectric constant of water vapour at 42 mc./sec., 99·8° C. and 76 cm. Hg, using a heterodyne method.

It had been intended to carry out work at other frequencies, temperatures and pressures, and with other gases, but this has been prevented by the war.

§ 5. ACKNOWLEDGEMENTS

This work has been carried out in the Natural Philosophy Department of the University of Melbourne, as part of the research programme of the Radio Research Board of the Australian Council for Scientific and Industrial Research. The author wishes to express his gratitude to Professor T. H. Laby, F.R.S., under whose direction the work was carried out, for his close interest and assistance at all times, and to Mr. R. W. Boswell, M.Sc., for many helpful suggestions.

REFERENCES

Tregigda, A. C., 1940. *Phys. Rev.* **57,** 294–297.
Hund, A., 1924. *Sci. Papers Bur. Stand., Wash.*, **19,** 487–540.
Megaw, E. C. S., 1936. *Wireless Engr.* **13,** 65–72, 135–146, 201–204.
Laby, T. H. and Hopper, V. D., 1939. *Nature, Lond.*, **143,** 240.

A DEMONSTRATION OF INDEPENDENT MEASUREMENT

By NORMAN CAMPBELL, Sc.D., F.Inst.P.,
London

MS. received 21 *October* 1942

ABSTRACT. Two and two sometimes, but not always, make four. That is the most important fact in physics ; for on it depend both the possibility and the limitations of measurement. Simple apparatus is here described by which the fact can be demonstrated to students.

1. Few, if any, courses of practical physics include any formal instruction in the principles of measurement. In particular no experiments are directed to demonstrating that the "fundamental" or "independent" measurement of a magnitude, involving addition and requiring no prior measurement of any other magnitude, is sometimes, but not always, possible. One reason for this neglect of a fact that lies at the root of all experimental physics may be the difficulty of choosing a suitable magnitude as an illustration. If length or mass (i.e., that which is measured by a balance and weights) were chosen, a demonstration that certain conditions have to be fulfilled would be unconvincing; for the student is already so familiar with the measurement of

these magnitudes that a suggestion that the necessary conditions should be violated would seem to him merely silly; he would not take seriously a suggestion that he should inquire what would happen if, in "adding" the lengths of two rods, he were to leave a gap between their juxtaposed ends. On the other hand, an attempt to illustrate the fact in one of the fields where the failure of addition leads to real difficulties (for example, heterochromatic photometry) would fail because the fact to which his attention is directed would be obscured by so many other facts with which he is equally unfamiliar.

This dilemma might perhaps be avoided by concealing from the student that the magnitude he is trying to measure is one familiar to him. Such concealment is certainly desirable initially, whatever magnitude is chosen; for one of the things that must be taught him is that independent measurement requires no prior knowledge of the magnitude, except that it is the field of a transitive asymmetrical relation; and nowadays even a young schoolboy cannot be trusted to have no prior knowledge about any magnitude. But this proposal does not entirely overcome the objections to choosing length or mass, or even electrical resistance. For it is not very easy to devise any experiment in which such magnitudes could be measured without disclosure of the operations that are being performed; and, even if it were possible, it would probably be educationally undesirable not to make the disclosure at a later stage; if the disclosure revealed something apparently silly, the lesson would be without effect.

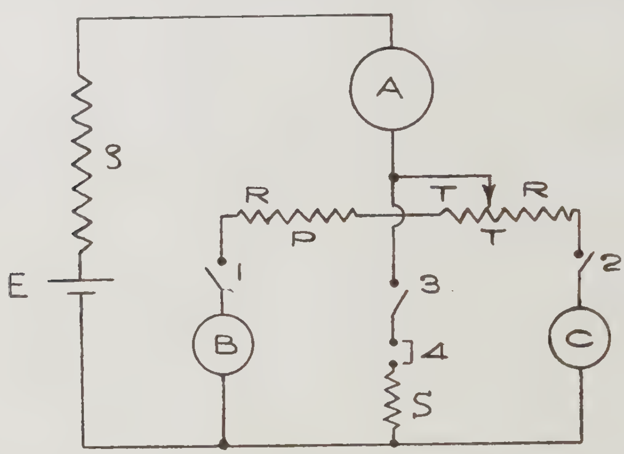
The purpose of this note is to suggest that electric current is a suitable magnitude for the demonstration. It is easy to conceal the operations performed; for the operating elements are switches and the function of a switch is not disclosed by its exterior form. Further, even if the fact that the operating elements are switches suggests that the magnitude concerned is current, the operations that they perform will not necessarily be guessed immediately or be judged immediately appropriate or inappropriate. For even a mature physicist, though he realizes at once, when his attention is drawn to the matter, that addition in branched circuits enables current to be measured independently, will sometimes be found not to have known that fact consciously; the possibility does not seem to be generally regarded as common sense.

2. Let us be clear at the outset what has to be demonstrated. The proportions $1+1=2$, $1+2=3$, etc., are definitions of 2, 3, ... By use of these definitions we can often (but not always) build up, from a given thing arbitrarily called 1, a "standard series", all different from each other and from the datum, to which the designations 2, 3, 4, ... must be assigned. The question whether in this series $2+2=4$, etc. is one of fact that cannot be predicted *a priori*, but must be established by experiment. If, but only if, $2+2=4$, the building up of the series leads to the kind of measurement useful in physics. (Of course it is not proposed to demonstrate this last statement; its proof requires a survey of all the uses of measurement.) In strictness, the trial should be carried further and the series extended so that the relations $2+3=5$, $2+4=3+3=6$, etc., are tested; and this should be pointed out to the student; but if he can be convinced of the significance of the $2+2=4$ test, the educational object will probably have been attained.

3. The apparatus proposed is shown in the figure below. A, B, C are three ammeters with unmarked scales. E, ρ represent means for varying the current through A; the current need not be D.C. as suggested in the figure; A.C. will serve. 1, 2, 3, 4 are contacts controlling the distribution among branch circuits of the current flowing through A. Contacts 1, 2, 3 are ganged in a multiple switch which has three positions:—(*d*) 1 closed, 2 and 3 open; (*e*) 2 closed, 1 and 3 open; (*f*) 1, 2, 3 all closed. Contact 4 is controlled separately.

First, contact 4 is always open. Then the current through A in position (*d*) passes wholly through B, in position (*e*) wholly through C, in position (*f*) is divided between B and C in a ratio determined by the setting up of the tapping T on the potentiometer resistor P.

The student is directed to proceed thus. He marks 1 some position of the pointer on the scale of A. In position (*d*) he brings the pointer of A to this mark by varying ρ and marks 1 on the scale of B the position of its pointer. In position (*e*) he similarly marks 1 on the scale of C. In position (*f*) he varies ρ and T so that both B and C stand at 1; he marks 2 the position of the pointer



of A. In positions (*d*) and (*e*) he marks 2 the corresponding positions on B and C. In position (*f*) he varies ρ and T so that B stands at 1 and C at 2; he marks 3 the position of A; and so on. He now in position (*f*) sets B and C both at 2 and notes whether A stands at 4. Owing to experimental error it will usually diverge slightly from 4; he therefore repeats the whole experiment until he convinces himself that, in the final test, A is as likely to stand on one side of 4 as on the other.

In the second half of the experiment, contact 4 is permanently closed. In positions (*d*) and (*e*) all the current still passes through B or C; this is the object of the contact 3. But in position (*f*) a fraction of the current, depending on the setting of T, passes through the shunt S and through neither B nor C. He repeats the operations precisely as before; but now in the final test the mark on A stands consistently below 4 by an amount that is greater than that which his previous observations have shown to be the experimental error. 2 and 2 do *not* make 4. Until he opens the apparatus, discovers what operations are performed by the multiple switch and the knob controlling the contact 4, and

interprets his discovery in the light of knowledge that he could not have possessed when the problem of measuring current *first* occurred, the difference between the two results will be a mere empirical fact; even if he were told that in one position of the knob 2 and 2 do make 4 and in the other not, he could not tell, without actual trial, which was which.

4. In order to show that the quantitative relations can be chosen suitably, the theory of the second experiment will be given. It will be assumed that the resistors obey Ohm's Law, but it must be insisted that this assumption is in no way necessary to the result of the first experiment; the measurement would be "true" even if all the resistors were non-linear or had intrinsic E.M.F.S.; all that is necessary is that there should be a one-one relation between the current passing through an ammeter and the position of the pointer.

In position (*f*) let the resistance between the centre of the potentiometer and the common terminal of B and C be R along each branch; let r be the resistance between T and this centre of the potentiometer. Let A, B, C be the currents through the ammeters and $\gamma = C/B$. Then from the circuit laws

$$\gamma = \frac{SR + r(S-R)}{S(R-r)} \qquad \ldots\ldots(1)$$

or

$$r = \frac{(\gamma-1)SR}{(\gamma+1)S+R} \qquad \ldots\ldots(2)$$

and

$$A = \frac{R(2S+R)}{S(R-r)}\,B = \left(\gamma + \frac{S+R}{S}\right)B. \qquad \ldots\ldots(3)$$

Let suffix x denote the value when the pointer of A stands at mark x in position (*f*). Then, in virtue of the process by which the marks are made,

$$C_{x+1} = A_x, \qquad \ldots\ldots(4)$$

$$B_x = 1. \qquad \ldots\ldots(5)$$

Therefore, from (3), (4), (5)

$$\gamma_{x+1} = \gamma_x + \frac{S+R}{S}, \qquad \ldots\ldots(6)$$

But $\gamma_2 = 1$. Therefore

$$\gamma_3 = \frac{2S+R}{S},$$

and generally

$$\gamma_x = \frac{(x-1)S + (x-2)R}{S} \quad (x > 1). \qquad \ldots\ldots(7)$$

and from (3), (5)

$$A_x = \gamma_x + \frac{S+R}{S} = \frac{xS + (x-1)R}{S} \quad (x > 1). \qquad \ldots\ldots(8)$$

If $S = 5R$, the currents corresponding to $x = 1, 2, 3, 4$ stand in the ratios 1, 2·2, 3·4, 4·6, so that $B_2 + C_2 = 4{\cdot}4$, while $A_4 = 4{\cdot}6$. If the error in a single setting of a pointer to its mark is δ, the maximum error in the reading of A in the test would be 17δ; the mean error would be about 6δ. δ need not exceed 0·01. Hence the difference between 4·4 and 4·6 should exceed the maximum error and be at least three times the mean error.

5. It is most desirable that the relation between current and deflection of the pointer should be different in each ammeter A, B, C, and that in two at least of them current and deflection should not be proportional; the more irregular the relation, the better. But an evenly divided scale of length may be provided, in order that the "marks" should be recorded in a notebook rather than imprinted on the instrument.

An intelligent student would guess the law of the "error" in the second experiment, if it was that represented by (8); he might suspect, therefore, that it is artificial, as indeed it is. This objection could be avoided by using for S (and possibly for the components of the potentiometer) a non-linear resistor.

It is desirable for ease of operation that the current through A should depend little on the setting of T. This, of course, can be achieved by making ρ sufficiently large compared with R or adopting some other constant-current device.

6. The apparatus can be used or readily adapted for two subsidiary demonstrations. One is the formation of sub-multiples of the unit. In principle this has to be done by finding n things all equal whose sum is 1; but if n is (say) 10, the process is impossibly laborious without some guide. In practice the necessary guide is usually derived from some algebraic law involving the magnitude, which can logically be established only after the measurement of the magnitude is fully established. But in our example the necessary law suggests itself immediately from the operations already performed. In the course of his proceedings, the intelligent student will have been led to suspect that if in position (f) he halves the reading of A, he will halve also the readings of B and C. That suggests that he can obtain the mark $\frac{1}{2}$, by setting A to 2, B and C each to 1 and then reducing A to 1 by changing ρ. An obvious procedure then leads to the mark $1\frac{1}{2}$. He can now confirm the suggested law, and the correctness of the method of subdivision, by ascertaining in position (f) that $1\frac{1}{2}+1\frac{1}{2}=3$.

7. Again, in extreme cases independent measurement fails at the outset because $1+1$ turns out to be 1; that is to say, combinations of similar things are similar to the components. Magnitudes of this kind are of great importance because they are "qualities", such as density. The possibility can be demonstrated by providing a fourth position (g) of the multiple switch in which A, B, C are all in series. Having marked 1 on B and C in positions (d) and (e), position (g) is adopted, and it is found that, when B and C each read 1, A also reads 1; no other numbers of the standard series can be established.

DEMONSTRATION

The interferometer in lens and prism manufacture. *Demonstrated to the Optical Group*, 17 *July* 1942, *by* F. Twyman, F.R.S.

A number of effects observable with the Twyman and Green Prism Interferometer were shown. They included a demonstration of the adjustment of the instrument, and examples of the distortion of wave front due to heating the

air or the optical medium through which the light passes, and the distortion due to the presence of vapours from volatile liquids. The optical effect of dropping warm water into cold was also shown, the effect in each case being made visible to the audience by the projection of the interference patterns on the screen.

The use of the instrument in optical manufacture was demonstrated by a series of lantern slides showing characteristic appearances on the interferometer, accompanied by a description of how errors are corrected by retouching. The use of the interferometer for testing the accuracy of angles was also described.

DISCUSSION

Mr. B. K. Johnson. As Mr. Twyman has mentioned the use of the interferometer for the testing of microscope objectives, I would like to ask whether the minute spherical convex mirror which is necessary in this test still consists of the globule of mercury originally suggested, or whether any other device is now used?

Also, has Mr. Twyman any knowledge of attempts being made by manufacturers of microscope objectives or others to correct the aberrations of such lenses by figuring processes, or is this still considered impossible owing to the smallness of the lens components?

Prof. A. F. C. Pollard. If Mr. Johnson will fuse a small bead on the end of a fibre of glass and silver it, I think he can easily obtain a perfect spherical mirror for high-power microscope objectives.

Some years ago I devised a glass total-immersion indicating hydrometer (*J. Sci. Instrum.* **1**, 1924, pp. 97 and 129) which rotated about a horizontal axis by the rolling of a small spherical glass foot on an optical flat. The error due to want of sphericity of the foot increases with the sensitivity of these little instruments, and for highly sensitive instruments the foot must be as perfect a sphere as possible. For a series of instruments I constructed, which indicated a change of the density of liquids to one in a million, the foot had a diameter of a little more than a millimetre. For such small diameters the motion of the fused glass due to surface tension greatly exceeds that due to gravity, and that part of the surface remote from the fibre-stalk settles into a nearly perfect sphere. There should be no difficulty in silvering the surface or depositing aluminium upon it. Such beads, sufficiently small for a 1/12th, could be mounted so that the objective can be tested in the interferometer in an immersed condition.

It has been reported that Leeuwenhoek constructed the minute lenses of some of his simple microscopes in this way, but as the cabinet containing 26 of his microscopes, which he bequeathed to the Royal Society, has unhappily disappeared, and no other specimens exist, we have no means of ascertaining whether this was the case or not.

The objection to Mr. Perry's suggestion of using bubbles in glass as spherical mirrors is that when a bubble is very near the surface of the glass it is pulled out of the spherical form. As the distance between the front surface of a 1/12th and its focus when immersed is about 0·2 mm. it would be difficult to find a bubble fulfilling the necessary conditions, but there is no difficulty in constructing perfect glass spheres less than 0·4 mm. in diameter.

Author's reply. In reply to Mr. Johnson, I am unable to say what is customarily used for the minute convex mirrors, but alternative devices are described in my paper on the microscope interferometer (*Trans. Opt. Soc.* **24,** 191 *et seq.*).

I do not know whether any attempts have been or are being made by manufacturers of microscope objectives or others to carry out correction of microscope objectives by retouching according to the indications of the interferometer. I do not think, however, that there would be any insuperable difficulty in doing so, since corrections could be carried out on the back lens, which is not too small for that purpose.

I am glad to know that Professor Pollard has found it possible to make a small bead on the end of a fibre of glass perfect enough to use as a back reflecting mirror for the interferometer, and I hope an opportunity may occur one day of trying such a bead for that purpose.

THE THEORY OF THE MICROSCOPE: III, BOUNDARY WAVES IN DARK-GROUND ILLUMINATION

By L. C. MARTIN,
Assistant-Professor of Technical Optics in the Imperial College of Science and Technology

Paper read to the Optical Group 22 January 1943

ABSTRACT. The theory of boundary diffraction waves (Young, Maey, Rubinowicz) is developed in a very simple way in order to deal with the theory of image formation in the dark-ground field, subject to the assumption of simplified conditions. Some confirmatory experiments are described. Attention is directed to certain anomalous results which may appear.

§ 1. INTRODUCTION

Two previous papers (Martin, 1931, 1934) on the theory of the microscope were concerned with a new analysis of the mode of formation of the image. In particular they sought to exhibit the relationship between the more familiar guiding principles, such as the "Abbe principle" and the "equivalence principle", which are commonly involved when the interpretation of the image is in question. For the most part, however, it was (as usual) assumed that the object was a very simple one, i.e. a grating with regular spacing, since the diffraction of light in such a case can be discussed in a comparatively straightforward manner with simple mathematical expressions. The second paper concluded with a note on the desirability of examining certain edge effects, encountered in the dark field, which had given anomalous results in experiments on the resolution of lines in a Grayson ruling. It seemed, therefore, appropriate to enquire into the dark-ground images of such objects as screen edges, both opaque and phase-retarding. In the course of a short study certain considerations emerged which are not well known and which have not, to the writer's knowledge, been applied to the theory of the microscope.

Diffraction by semi-infinite screens of various assumed properties is a subject to which a great deal of theoretical attention has been paid. In these problems the surface integrals familiar in text-book diffraction theory (Huygens, Helmholtz, Kirchhoff) can be replaced under certain conditions by line integrals along the boundaries of diffracting screens.* The pioneer paper was due to Rubinowicz (1917). This is concerned with the mathematical expression of the phenomenon of the luminous line, apparently at the diffracting edge, seen from a view-point within the shadow. Although physics text-books invariably give much attention to the distribution of light in a plane behind the diffracting object, the writer knows of only one † in which this important effect is discussed,

* A useful mathematical review is given by Baker and Copson in *Huygens' Principle* (Oxford University Press, 1939).

† R. W. Wood, *Physical Optics*, 3rd edition.

although it was explained much earlier by Maey (1893) on somewhat different lines, and has been investigated by a number of experimenters. The boundary-wave conception dates back, indeed, to the ideas of Thomas Young. Jentsch (1910) has given a survey of both theoretical and practical investigations. in which he comments upon the importance of the phenomenon in relation to dark-ground illumination in the microscope.

In the present paper a brief explanation of the phenomenon will be given in terms of the familiar Cornu spiral, the problem being treated as two-dimensional. Referring to figure 1, let O be the source at a distance a from

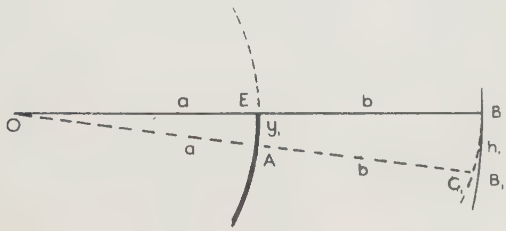


Figure 1.

the edge E of the thin opaque diffracting screen represented by the thick line. The wave front centred in O is shown by the broken line. Let the diffraction effects be observed in a locus BB_1, also centred in O; the distance EB is b. The point B_1 is situated at a small distance h_1 within the geometrical shadow. so that the edge E is at a height y above the straight line OB_1. In the elementary theory it is assumed that y is so small that quantities depending on higher powers of y than the second can be neglected. It is then easily shown that the excess of the straight line EB_1 over EB is given by $(a+b)y^2/2ab$ within the given approximation.

Putting $\dfrac{\pi v^2}{2} = \dfrac{2\pi}{\lambda}\left(\dfrac{a+b}{2ab}\right)y^2$, where λ is the wave-length of light, the amplitude at a point in the locus BB_1 is expressed conventionally in terms of the Fresnel integrals as $C+iS$, where

$$C = \int \cos\frac{\pi v^2}{2}\,dv,$$

$$S = \int \sin\frac{\pi v^2}{2}\,dv,$$

and the limits of integration correspond to the exposed part of the wave front. In the Cornu spiral, the numerical values of C and S obtained by taking zero for the lower limit and increasing values of v for the upper are plotted as the X and Y co-ordinates of a curve, the properties of which are fully explained in the text-books.

It is found that:—

Length of curve from O to v = v,
Angle of slope in radians = $\pi v^2/2 = \phi$ (say),
Curvature at the point v = πv.

We note that ϕ is the relative phase angle of retardation for a disturbance which is derived from an element of the wave front characterized by v as compared with an imaginary one which travels along a straight route ($v=0$) from the source to the point of observation; we also note that

$$v=\sqrt{\frac{2\phi}{\pi}}=y\sqrt{\frac{2(a+b)}{\lambda ab}}.$$

Boundary wave

We now consider the light from a point source, diffracted within the geometrical shadow of the straight-edged screen. The projection y_1 of the edge of the screen above the straight line OB_1 is given by

$$y_1=\frac{h_1a}{a+b}.$$

If B_2 is another point at a distance h_2 below B, the similar projection is

$$y_2=\frac{h_2a}{a+b}.$$

At both y_1 and y_2, part of the upper part of the wave surface is cut off, and the corresponding values of v in the Fresnel integrals giving the amplitudes at the points are

$$v_{1,2}=h_{1\,2}\sqrt{\frac{2}{\lambda}\cdot\frac{a}{b(a+b)}}.$$

Now, referring to the Cornu spiral diagram (figure 2), suppose that the values v_1, v_2 correspond to the points G_1, G_2 respectively; then at the points B_1, B_2 the effect (in phase and amplitude) of the whole upper part of the wave

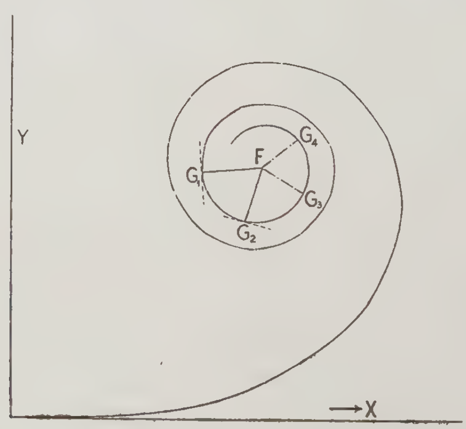


Figure 2.

will be represented (within the limits of accuracy already discussed) by the vectors G_1F and G_2F respectively.

Owing to the close approximation of the spiral curve to the circular form, the angle between G_1F and G_2F will be very nearly equal to the angle between

the tangents to the curve at G_1 and G_2 respectively. (This will only apply with reasonable accuracy for positions of G_1 and G_2 with corresponding v values exceeding $\sqrt{2}$.) But the angle of slope for the tangent is given by $\pi v^2/2$. Consequently the difference of the phase angles of the resultants G_1F and G_2F is given by

$$\frac{\pi}{2}\left(v_2^{\,2}-v_1^{\,2}\right)=\frac{2\pi}{\lambda}\,\frac{h_2^{\,2}-h_1^{\,2}}{2b}\left(\frac{a}{a+b}\right). \qquad \ldots\ldots(1)$$

The greater value of h corresponds to the greater path, and consequently to greater retardation.

If we draw a new surface BC_1 (figure 1) concentric with the edges of the screen, the distance between this and the surface of observation at any distance h below the geometrical shadow edge will be

$$\frac{h^2}{2b}-\frac{h^2}{2(a+b)}$$

if we neglect quantities depending on h^4 and higher powers of h. The corresponding phase difference is

$$\frac{2\pi}{\lambda}\cdot\frac{h^2}{2}\left(\frac{1}{b}-\frac{1}{a+b}\right)=\frac{2\pi}{\lambda}\cdot\frac{h^2}{2b}\left(\frac{a}{a+b}\right),$$

and, therefore, for the two distances h_1 and h_2 we have an increase of the phase difference (corresponding to the increase of the gap between the curves)

$$\frac{2\pi}{\lambda}\cdot\frac{h_2^{\,2}-h_1^{\,2}}{2b}\left(\frac{a}{a+b}\right),$$

which is the same as the value in (1) above. The greater the value of h is this new surface, the less is the straight path from the source O to the new point considered. Consequently, although the resultant of the upper part of the wave shows a progressive phase change in surface BB_1, there will only be a negligible phase change in BC_1; in other words, *the resultant of the total upper part of the wave behaves as if it were spreading out from a source at E.*

Consequently a lens or an eye with a relatively small aperture placed so that it can only receive light diffracted well within the geometrical shadow edge will give an image of a bright source apparently coincident with the edge of the diffracting screen. The difference from a real source is found in the diminishing amplitude of the disturbance with increasing distance within the shadow. This, however, will produce an appreciable effect on the image only when the numerical aperture of the observing system is large. For a small numerical aperture, the apparent brightness of the image will be proportional to the square of the length of the appropriate vector line GF. The diminution of the amplitude is, of course, rapid at first, but then slower, so that appreciable light is diffracted to large angles within the "dark" region. The treatment is valid if the source O is a line source parallel to the edge E. In this case the edge appears illuminated continuously as a bright line.

We will now consider the effect at B_1 of a bar or rectilinear obstacle MN placed above the edge E (figure 3), so that its edges M and N are parallel to E. The heights of M and N above OB_1 correspond respectively (say) to v_3 and v_4.

The effect of the wave from E to M is represented in the spiral diagram by G_1G_3, and from N upwards by G_4F. Consequently the whole effect at B_1 is represented (on the Kirchhoff theory) in amplitude and phase by the vector sum of these components.

Now as to the effects observed from the region of B_1. It is clear that in this region the displacement G_1G_3 can be compounded from G_1F *plus* FG_3 or G_1F *minus* G_3F. If these last displacements be discussed on the lines above, it can be shown in a similar way that they correspond to boundary waves centred in E and M respectively; moreover, the effect near B_1 of the wave above N can be summarized in a boundary wave (G_4F) centred in N. The conditions are, therefore, consistent with the following phenomena observable at B. The edges of the screen and parallel diffracting obstacle are seen illuminated with an intensity corresponding to the squares of the corresponding lengths G_1F, $-G_3F$, and G_4F; moreover, the relative phases of the resulting disturbances are represented by the relative angles of these lines in the Cornu diagram.

Consider, for example, that the diffracting bar MN decreases continually in breadth; the corresponding points G_3 and G_4 move closer and closer together.

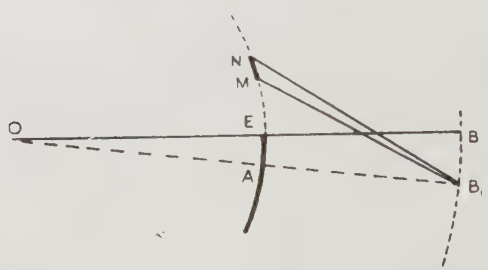


Figure 3.

As observed from B_1, the boundary haloes will remain little affected in appearance till the bar is so narrow that the angular separation of M and N at the observing point is comparable with the resolving limit of the observing instrument. As the displacements (characterized by $-G_3F$ and G_4F) of the disturbances reach numerical equality their phase difference approaches π; and the boundary waves naturally extinguish each other as the breadth MN vanishes.

It may be inferred that in order to get appreciable light diffracted from a *very narrow* obstacle it will be profitable to make the equivalent y as large as possible; the edge waves may then show an appreciable departure from the phase difference of π characteristic of vanishing thickness; this should apply to the case of the microscope—a big aperture of the observing system conducing to high values of y (very large angles of diffraction).

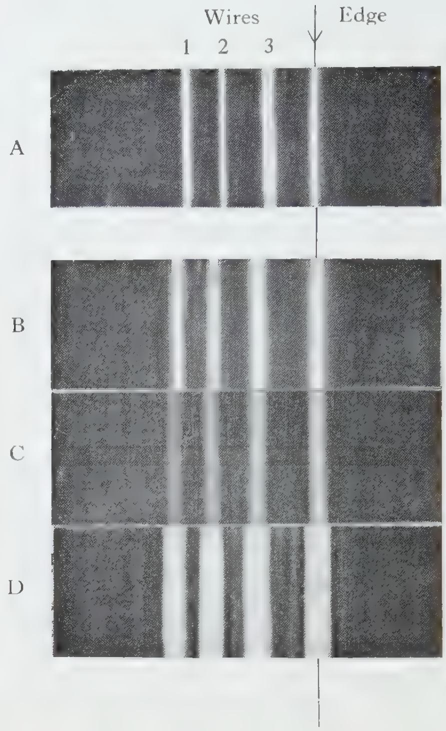
In the same way it may be argued that the boundary waves from M and E will have a phase difference of π when M and E approach very closely together.

§ 2. EXPERIMENTAL OBSERVATIONS

An optical bench was set up with a spectrometer slit (illuminated by an arc lamp) as a source of light, a safety-razor blade to serve as an opaque straight-edged screen, and an observing telescope with a rectangular entrance pupil

Plate 1.

Diffracted boundary light at a razor edge.

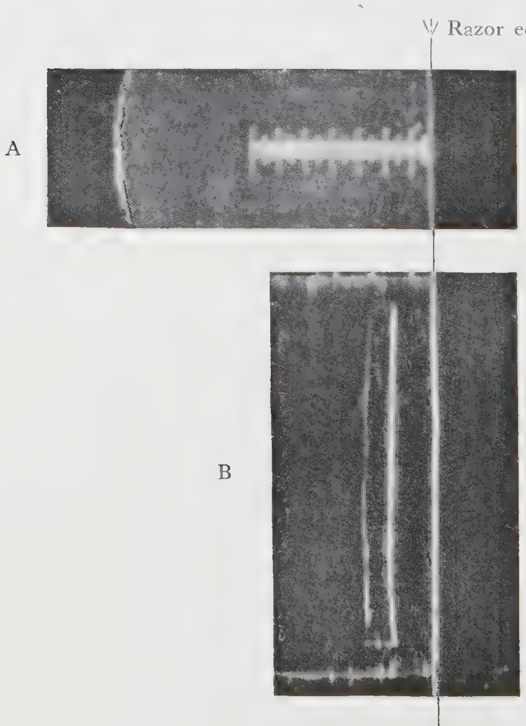


	Approximate resolving limit (mm.)	h_1 (razor edge)
A	0·2	11·25 mm.
B	0·18 (long exposure)	14·5 mm.
C	0·9	7·5 mm.
D	0·9 (long exposure)	8·5 mm.

Plate 2.

Diffracted boundary light from wires.

Wire 1, thickness = 0·118 mm.
,, 2, ,, = 0·018 ,,
,, 3, ,, = 0·227 ,,



	Resolving limit	h_1 for razor edge
A	0·2 mm.	11·25 mm.
B	0·18 mm.	15·5 mm.

Plate 3.

A—Diffracted light from graticule graduations ; breadth of lines = 0·009 mm. approx.
B—Diffracted light from edges of shallow etched groove in glass plate.

To face page 109

of which the width in the y direction perpendicular to the diffracting edge could be controlled from zero up to 2·5 cm. A convenient focal distance for the telescope was found to be 28·5 cm. The eye-piece could be removed and a plate-holder substituted. The distances a and b were both 1 metre. The image of the diffracting edge could thus be photographed; the final magnification in the plates reproduced herewith is 3·4. The values of v corresponding to $\phi=0, 2\pi, 4\pi, 6\pi$, etc., are $0, \sqrt{4}, \sqrt{8}, \sqrt{12}$, etc., and the y values corresponding to successive complete turns of the spiral are 0·707, 1·0, 1·224, 1·414 mm., so that the intervals corresponding to the first four complete turns are 0·707, 0·293, 0·224 and 0·190 mm. The elements get rapidly narrower, and the following short table gives the width of half-period elements for some larger values of y.

Widths of half-period elements for $\lambda=0{\cdot}5\times10^{-4}$ cm.

y (mm.)	Width (mm.)
4	0·031
5	0·025
6	0·021
7	0·018
8	0·016
9	0·014
10	0·012

For blue light the elements are narrower and for red light broader than for the mean wave-length.

Now, for an entrance pupil of 2 mm., the numerical aperture (NA) of the observing instrument is 10^{-3}, so that its expected resolving limit in the object plane will be given by

$$\frac{0{\cdot}5\lambda}{\mathrm{NA}} = 0{\cdot}25 \text{ mm.}$$

The first photograph, plate 1, shows the boundary wave from the safety-razor edge photographed with NA = 0·00125, the centre of the entrance pupil being 11·25 mm. within the shadow. The edge is seen to give an effect indistinguishable from a linear source of light. The lateral fringes due to diffraction by the aperture of the observing system can be seen in the original negative.

Plate 2 shows four photographs, in all of which the boundary light of the razor edge appears on the right. Three fine wires were supported in the light to the left of the edge with the aid of a small frame, and their "dark-ground" images are seen. A was taken with a position of the frame differing a little from its location in B, C and D. Wires 1 and 3 were of copper (diameters 0·118 and 0·227 mm. respectively); they were much tarnished. Wire 2 was of tungsten (0·018 mm.) and very highly reflecting. A was taken with white light, but using an "ordinary" plate; only the thickest wire (3) is resolved by the separation of the boundary light from its edge, and the performance in resolution is normal; but it was at first surprising to notice how nearly equal in intensity were the resultant wave from 2, the two separated edge waves from 3, and the single wave from the razor edge. The visual effect of wire 1 was not very dissimilar—but it is approaching resolution and looks broader in the photograph. Photograph B was actually taken with an aperture exceeding that for A, and was also

heavily exposed, so that any lack of symmetry in the lateral fringes might be detected; but only very slight effects of this kind can be noticed.

In C the aperture of the observing system was made very small so that the resolving limit is about four times the diameter of the thickest wire. An effort was made to choose the exposure so that the densities of all four images might fall on the straight part of the photographic plate characteristic, and thus represent by their relative brightness something approaching the true original effect. It is then seen that these dark-ground images show a simple progression of brightness from the razor edge outwards, just as if the amount of light obtainable from any such object (below the resolving limit) were more or less independent of its size (within a considerable range) or reflecting power, etc., and were, in fact, not greatly different from the effect of the razor edge (i.e. amplitude $\equiv$ GF as above).

Photograph D differs only from C in having a much larger exposure; the lateral diffraction maxima are seen and the apparent width of the lines has increased with over-exposure.

The elementary theory given above relates to supposed thin screens of complete opacity. The Huygens-Kirchhoff procedure of integration over the exposed parts of the wave-front is usually found to be justified by its results, but the thin wires used in the present experiments differ in important respects from the "screens", since reflection can take place at the near side of the wire and enter the aperture of the observing system. The diffraction of light by a reflecting cylinder has already been studied on the basis of the electromagnetic theory, and the results of such a theory would have to be taken into account in any experiments of a more exact character. But the observations show that some of the broader effects can be accounted for by the above elementary theory, in which the reflection is neglected in regions near the shadow boundary.

In support of this view, experiments were made in which the fine lines in photographic (platinum) graticules were used in place of wires. These have a thickness presumably much smaller than any of the wires, so that the amount of light reflected into the shadow from their edges would be much smaller, but the boundary-wave phenomena were very similar. Such a line is about 0·009 mm. wide. A graticule of this kind (1 cm. divided into 100 parts) was placed just behind the razor edge and partly screened by it. Every fifth line is longer than the rest. The resolving limit was about 0·2 mm., so that the shorter lines (0·1 mm. apart) are not resolved from each other, but the longer lines appear (plate 3, A) to be illuminated with very closely the same brightness as the edge. Now the value of y corresponding to the edge in this photograph is 5·625, so that the width of one half-period element is 0·023 mm. for yellow-green light (probably rather smaller for the effective wave-length in the photograph). The breadth of the line corresponds, therefore, only to about 70° in the Cornu spiral. It might be expected that anything much finer than this would begin to diffract less light than the edge itself.

Colour effects

If the observing NA does not exceed 10^{-3}, the boundary waves F_3 Gand G_4F from the edges of an obstacle smaller than about 0·2 mm. will interfere. For blue light the corresponding phase-angle difference will be greater, and for red

light smaller, than for the mean. According to the circumstances, then, there may be a wave-length or wave-lengths within the visible spectrum for which the phase difference of the disturbances interfering in the image is π, and as the vectors are not appreciably different in length, such a colour will be absent in the combined result. The two white boundary waves seen in the field with a larger aperture and with white light now merge into one *coloured* line; coloured images were seen by the writer when the observing pupil was very near the edge of the shadow. Imagining a linear opaque obstacle in the field to increase in thickness from zero upwards while the effect is watched, it is clear that the phase difference of the boundary waves will (starting from π) vary most rapidly for violet light, and consequently we should expect the colour to start from a dark blue-violet and go through a sequence of "interference" colours similar to those of Newton's rings.

If (in contrast to the above case) a wire of fixed diameter is observed in a narrow-slit system, and moved so that the y value of the wire increases, it is clear that for any given wave-length the phase angle between the boundary waves must also increase, since the v interval for one turn continually diminishes as we approach the centre F: and the rate of change is quickest for blue light; consequently the colour changes with the movement of the wire. In the case of the broad slits, the combined colour effects of the superposed systems combine to neutralize each other, and the residual effect is white.

With an observing slit of finite width and/or an illuminating slit not negligible in size, we have, in effect, superposed the combined effects of a number of thin-slit systems with different y values of the centre of the wire.

Thus, if broad slits are used it will be understood that while (sets of) two boundary waves are superposed in the unresolved image of the wire, the intensity, instead of being double that of either edge halo, is weakened by much interference, whereas the razor-edge boundary waves with broad slits are additive. The nett result is that the single image of the wire does not differ greatly in brightness from the boundary light of the razor edge. Moreover, the effect with wires of different thickness is a complex one, and the image of the thickest wire is not necessarily the brightest. Small changes in the y value with broad slits do not provide much apparent change in the brightness of the unresolved image. Whether a wire is seen resolved with two boundary waves or unresolved "as a single wave", the lines of bordering light appear to casual inspection to differ little in brightness from each other. It would doubtless be possible to develop this point theoretically and numerically, but the diffraction conditions characteristic of any material obstacle are so complex that it is a task of some magnitude.

Phase-contrast object

As distinct from an opaque linear obstacle, an important type is, of course, a phase-changing strip of material. Cylindrical objects are common in the microscope, but a flat ribbon of material makes a case for which a simple theory can be given in terms of the Cornu spiral. Although R. W. Wood (*Text-book of Physical Optics*) used that diagram to discuss the distribution of intensity in the region of diffraction from the straight edge of a phase-retarding lamina, the specific problem of the boundary wave at the edge does not appear to have

been discussed by this means, though related effects as observed in the conditions of the Foucault test have been discussed mathematically and numerically by Ghosh (1919). We will take the case of a straight-edged lamina first.

If a retarding plate has the shape suggested in figure 4, a wave having passed through it can be considered to consist of two separate parts, EA and BC, together with a connecting part AB characterized by a progressive change of optical path. The width of AB depends entirely on the characteristics of the retarding plate, and unless this width is small in comparison with a half-period zone, the transition region produces a marked effect on the phenomenon.

As explained above, the edge-wave from E found within the shadow is represented in phase and amplitude by such a vector line as would join the points E and F (figure 5). The edge-wave at A might be FA. If the width AB (figure 4) were negligible and the retardation also negligible, the edge-wave from B would be AF (cancelling FA). But if the width is negligible while the retardation is finite the case will be represented by adding to FA the vector obtained by rotating

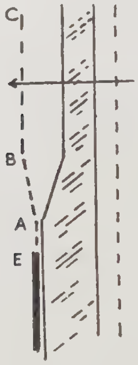


Figure 4.

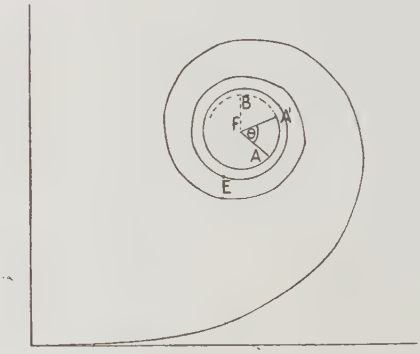


Figure 5.

AF through the appropriate phase angle suggested by θ in the figure. The resultant is the vector sum of FA and AF. Since the phase angle varies with the wave-length, the resultant amplitude also varies with wave-length, and the combined boundary waves formed with incident white light will be visibly coloured unless the retardation amounts to several complete wave-lengths and there are a number of absorption bands distributed through the spectrum.

Experiments with a phase-contrast object

The object used by the writer was a small plate of optically worked glass in which a groove about 1 mm. wide and one wave-length deep had been etched by means of dilute hydrofluoric acid. The edges of the groove were arranged parallel to the knife-edge and their appearance was observed from an aperture placed within the shadow. We thus have one step of diminishing, and one step of increasing path. Strongly coloured edges were seen, as expected, when the aperture was very close to the shadow edge—but when the y value had increased to about 8·8 mm. for the centre of the aperture (NA $= 0{\cdot}0015$) the edges both

appeared white. In each case the nearer edge of the groove (where the optical path diminished with increasing distance from the knife-edge) was much brighter than the other, and this brighter edge differed inappreciably from the edge-light from the razor blade (see plate 3, B).

A much narrower groove in the same plate (below the limit of resolution) appeared as a single white line, again of about the same brightness as the knife-edge. This is not shown in the photograph.

It is evident from the inspection that edges of the wider groove are "shelving" instead of clear-cut, and the region of transition probably occupies the width of several half-period elements; moreover the total path difference changes with wave-length. The disappearance of colour with incident white light and slit widths of appreciable size is therefore understandable, and the greater intensity of the near-side edge is explainable as a refraction effect. However, a closer analysis of the edge-effect with the aid of the Cornu diagram is instructive. Take the case of one edge first.

§ 3. THEORY OF EDGE EFFECTS

Let AB (figure 4) be the region of transition; then suppose the edge-wave at A is FA (figure 5). Suppose for a moment that there is no step in phase, then the transition region is AB (taken along the curve in figure 5) and the edge-

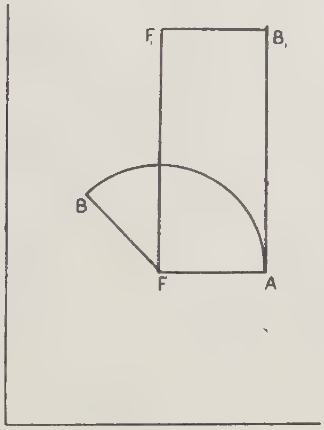


Figure 6.

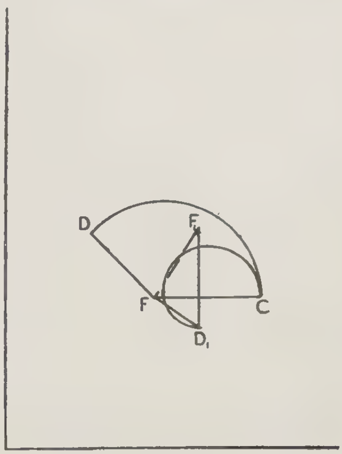


Figure 7.

wave at B is BF. If the transition region is unresolved, the whole wave is represented by the vector sum of these lines, and since

$$FA + AB + BF = 0,$$

no light is seen if there is no phase-step. If, on the other hand, a change of phase takes place, there is a marked difference between the effects of an increase or diminution of the path. On the near side of the groove there is a diminution of the path corresponding to a clockwise turn in the diagram. In the simplest case, all the elements of AB would experience a clockwise turn proportional to their distance from A. Suppose, for a concrete example, the diminution in path between A and B corresponds to a phase angle of 135°, while the arc AB

in the Cornu spiral is also 135° (see figure 6). The effect is to straighten out the curve into the line AB_1. We have now to add the vector obtained by rotating BF through 135° clockwise, thus obtaining B_1F_1; the resultant is FF_1.

Compare with this the effects on the far side of the groove where there is an increase of path; let us suppose this to be of the same amount (figure 7), and let the region of transition CD be equally broad. Each element of CD will now have to rotate anti-clockwise, and the spiral will curl up at twice its original rate, taking the form CD_1. On adding DF to this, after rotation anti-clockwise through 135°, the resultant is FF_1. The tendency for the spiral to uncurl in one case and to curl up in the other readily explains the greater average length of the amplitude vector in the case of diminishing path above.

Effect of broad slits

The "source" slit can, in cases similar to those discussed above, be considered to be resolved into a series of imaginary elementary slits A_1, A_2, A_3, etc., each of which gives rise to boundary waves centred in the edges E, M, N, etc., of diffracting obstacles. If the aperture of the observing system is large enough to "resolve" these edges, the intensities due to the elementary slits are simply additive, since A_1, A_2, A_3, etc., give incoherent light, but as soon as the aperture is small enough to prevent resolution of the edges there will be interference, depending on the relative phases of the supposed pairs of coherent border-waves—and these relative phases will vary to some extent with the position of the elementary slit. If we know these relative phases and the numerical aperture of the observing system, we can calculate the light intensity in the final image produced by this system for every elementary source, and integrate to get the total effect. The assumption is, of course, that the edges act as the line sources of coherent light, which is nearly correct for small observing apertures, but would have to be modified in the case of the large apertures on account of the continual shrinking of the light vector with increasing y value. Moreover, the present treatment at least has not proved that the geometrical edge of the obstacle will be the *exact* centre of phase constancy when the angle of diffraction is large, and obstacles of finite thickness present complex problems.

§ 4. APPLICATION TO MICROSCOPICAL OBSERVATIONS

It is very remarkable that in spite of all the attention which has been given to the technical side of dark-ground illumination in the microscope, and the theoretical and experimental controversies which have taken place regarding it, the theory of boundary waves does not seem to have been applied to the question of the image formation of the small bodies such as bacterial cells and filaments; their illuminated boundaries are the most obvious phenomenon in a dark-ground image. But once the conception is formed it gives a *principle* by which the mode of formation of the image formation of such single objects can be described; and this principle is of a character similar to Abbe's principle for discussing the image-formation of periodic structure. Moreover, although its most obvious application is to dark-ground work, the boundary-wave conception may be applied even in the bright-field case; the discussion of this point must, however, be postponed.

To formulate the principle fully it will clearly be necessary to consider the more general expressions derived by Rubinowicz and others for the phenomena in which the two-dimensional simplicity of the above elementary discussion is necessarily dispensed with; but even at the present stage it points to some considerations of great importance in practical high-power microscopy.

For example, in using a very narrow but high-aperture illuminating beam in dark-ground microscopy with one-sided illumination it is (on the above lines) possible to formulate conditions under which the boundary waves from an opaque strip (derived from one elementary source associated with the beam) have opposite phases as referred to the centre of the observing aperture. Suppose that the object has a diameter which is half that represented as resolvable by the usual formula. The overlapping of the corresponding diffraction patterns in the field of the observing instrument occurs as shown in figure 8, in which

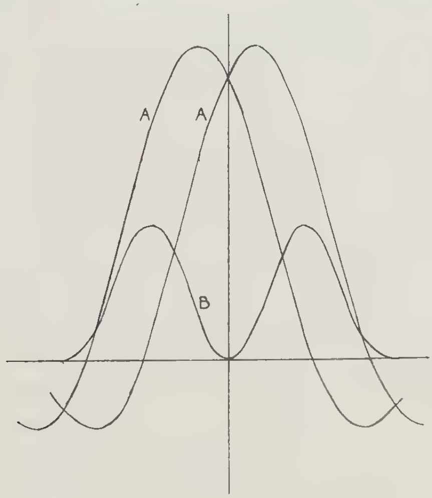


Figure 8.

A. Relative amplitude of displacement in image.
B. Relative intensity of resultant if superposed images have opposite phases.

the separated amplitudes are plotted, as also the resultant intensity, B. The resultant intensity maxima are separated by a gap of perfect blackness, and they are nearly twice as far apart as the original amplitude maxima.

The numerical aperture-range of the illumination for an elementary "slit" source would be infinitesimal, but the relative phases of the boundary waves need not change greatly for a finite range of illuminating aperture, and results similar to the above might still be possible. Even with symmetrical dark-ground illumination they may still persist, especially with monochromatic light.

Since the writer's previous papers on the theory of the microscope indicated that a true equivalence to self-luminosity is only attained when the numerical aperture of illumination is unrestricted, the use of dark-ground illumination must, in effect, give a large measure of coherence in the illumination; and if this is the case, the use of the boundary-wave principle indicates at once that the "image" of an object having a size near the limit of resolution may bear an

unexpected geometrical relation to the object, while spurious "resolutions" may occur. The writer does not wish to claim that the "theoretically-forbidden" performances in resolution reported by Merling-Eisenberg (1937) and others are hereby explained, but that the boundary-wave principle offers a means for the closer theoretical discussion of such dark-ground images.

Again, if the boundary waves are in a large measure coherent, the application of the Gerhardt (1926, 1927) method (the adaptation of the Michelson stellar measurement interference method to the measurement of sub-microscopical particles) ought certainly to take this condition into account instead of allowing the assumption that the light at the two sides of the boundary is simply incoherent; otherwise very misleading results may be obtained.

It is hoped to follow up the present paper by further discussion and experiment with a view to a closer understanding of the results of dark-ground illumination.

REFERENCES

Gerhardt, U., 1926. *Z. Phys.* **35,** 697 ; 1927. **44,** 397.
Ghosh, P. N., 1919. *Proc. Roy. Soc.* A, **96,** 257.
Jentsch, 1910. *Handb. Phys.* **21,** 910.
Maey, E., 1893. *Ann. Phys., Lpz.,* **49,** 69.
Martin, L. C., 1931. *Proc. Phys. Soc.* **43,** 186 ; 1934. **46,** 233.
Merling-Eisenberg, K. B., 1937. *J. Quekett Micr. Club,* **1,** 311.
Rubinowicz, A., 1917. *Ann. Phys., Lpz.,* **53,** 257.

THE AIRY DISC FORMULA FOR SYSTEMS OF HIGH RELATIVE APERTURE

By H. H. HOPKINS, B.Sc., A.Inst.P.,
Messrs. Taylor, Taylor & Hobson, Ltd.

MS. received 26 *October* 1942; *read to the Optical Group* 5 *January* 1943

ABSTRACT. The light distribution near the focus of high-aperture systems is investigated, account being taken of two factors which are neglected in the derivation of the Airy disc formula for low-aperture systems. These are : (*a*) an inhomogeneity of amplitude which is shown to be present over the emergent spherical wave-surface, and (*b*) the fact that light vectors at the focus of the wave are perpendicular to the ray-paths associated with them, it being shown that it is only necessary to consider the components of such light vectors in the direction of the light vector of the incident polarized wave. A further consideration is that it is not necessary to assume that the emergent spherical wave-surface is coincident with the plane of the exit pupil.

The inhomogeneity of amplitude is shown to be different for the case of a Gaussian system and for one obeying the sine condition : the problem is solved for each of these cases, and conclusions are drawn concerning the limits within which the low-aperture formula is valid, together with some remarks on the resolving power of high-aperture systems.

The Airy disc formula was derived to apply specifically to the light distribution in the geometrical focal-plane of a telescope objective. With the usual Bessel function notation, it takes the form

$$I=\{J_1(x)/x\}^2, \qquad \ldots\ldots(1)$$

in which I is the intensity at a distance ρ' off axis, and $x=\frac{2\pi}{\lambda}\cdot\frac{R}{F}\cdot\rho'$; R is the radius of the exit pupil and F the focal length of the objective. If x_0 is the first zero of the function $J_1(x)/x$ and α the semi-angle of the light cone proceeding from the objective, then according to the formula as quoted,

$$d=\frac{\lambda}{2\pi}\cdot\frac{1}{\tan\alpha}\cdot x_0,$$

where d is the least resolvable distance according to the Rayleigh criterion. This gives

$$d=0{\cdot}61\,\frac{\lambda}{\tan\alpha}. \qquad \ldots\ldots(2)$$

In these expressions $d\to 0$ as $\alpha\to\pi/2$, a conclusion which suggests that the formula is not valid for larger angles. However, the approximations in the derivation of the Airy disc formula are equivalent to setting $\cos\alpha=1$, so that in these expressions one can write equally $\sin\alpha$ or $\tan\alpha$. If the former is employed, the results for $\alpha\to\pi/2$ are still intelligible; and in fact, although not stressed, the use of the variable $x=\frac{2\pi}{\lambda}.\sin\alpha.\rho'$ is implied in a paper by Silberstein (*Phil. Mag.* **35**, 1918), in which integrals expressing the effects of the presence of spherical aberration on the Airy-disc distribution are discussed by a method analogous to that employed in Cornu's spiral. Except for this reference, however, the elimination of the approximations employed in Airy's calculation does not appear to have been considered. It is the purpose of this communication to derive formulae applicable to systems of high relative aperture.

In the first place, let the problem to be considered be that of the imagery of an infinitely distant axial point-source by an ideal Gaussian system. A second problem is discussed relating to systems which fulfil the sine condition.

§ 1. GAUSSIAN SYSTEMS

In figure 1, $H_1y_1z_1$ and $H_2y_2z_2$ are the unit planes at right angles to the optic axis H_1H_2 of an ideal Gaussian system. The periphery $a_1b_1c_1d_1$ of the small area S_1 in the plane $H_1y_1z_1$ is imaged point for point in the periphery $a_2b_2c_2d_2$ of the small area S_2 in the plane $H_2y_2z_2$. Apart from negligible diffraction effects, all the energy passing through S_1 will also pass through S_2, and since the areas of S_1 and S_2 are exactly equal, the amplitude of a wave emerging at S_2 will be equal to that of the wave incident at S_1. Thus, assuming complete transmission at each surface in the system, the elements of an emergent spherical wave will have equal amplitudes as they cross the plane $H_2y_2z_2$, providing the incident wave be both uniform and plane.

If points on the periphery of S_2 be joined to the focus O (figure 2), the resulting lines are emergent ray-points, and if they intersect the spherical emergent wave-front at S_3, the ratio of the amplitudes of the disturbances at S_3 and S_2 will be equal to

$$\left(\frac{\text{area } S_2}{\text{area } S_3}\right)^{\frac{1}{2}}.$$

Thus. as the element S_3 is taken further off the axis, the amplitude over the spherical cap of the emergent wave-front increases. It is of importance in formulating the Airy disc formula for high apertures to take account of this inhomogeneity of amplitude, and it is in this respect that a system fulfilling the sine condition differs from that now considered.

In figure 2 YOZ is the paraxial focal plane and O the paraxial focus. The amplitude of the emergent wave is constant over the unit plane $y_2H_2z_2$, and

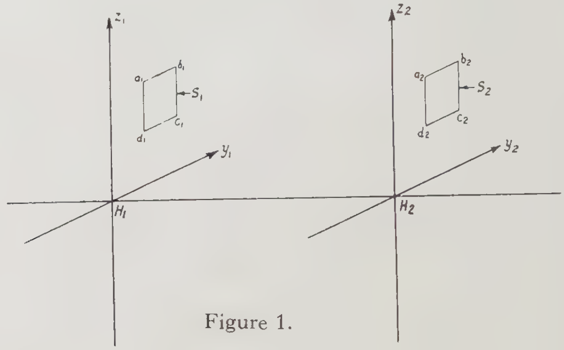


Figure 1.

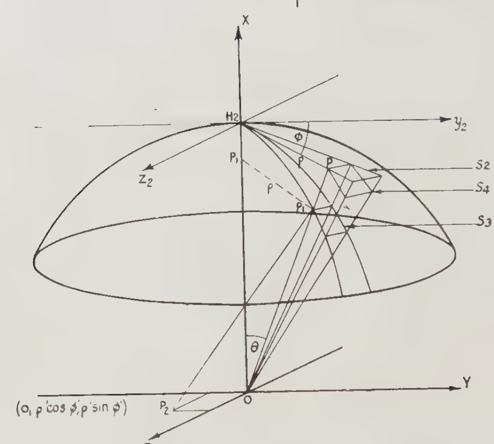


Figure 2.

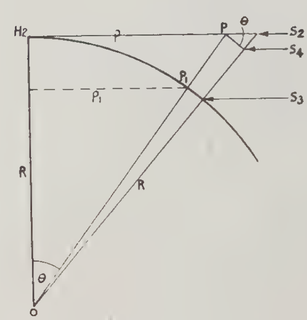


Figure 2 *a*.

differs for different annuli over the spherical emergent wave-front. The element of area S_4 is parallel with the element S_3. The ratio

$$\frac{\text{area } S_4}{\text{area } S_3} = \frac{OH_2{}^2}{OP_1{}^2 \cos^2\theta} = \frac{1}{\cos^2\theta}$$

and

$$\frac{\text{area } S_2}{\text{area } S_4} = \frac{1}{\cos\theta},$$

so that

$$\frac{\text{area } S_2}{\text{area } S_3} = (\cos\theta)^{-3}, \qquad \ldots\ldots(3)$$

and if the amplitude is unity on the axis and over the plane $y_2H_2z_2$, it is $(\cos\theta)^{-3/2}$ over an annulus whose angular co-ordinate is θ.

Suppose the incident wave to be plane polarized in the plane XOZ. The emergent wave is then characterized by the fact that the light vector is both perpendicular to the emergent ray-paths and contained in planes such as $P_1N_1OV_1$ and $P_2N_2OV_2$ (see figure 3). $P_1'H_2P_2'$ is the line of intersection of the emergent wave-front with the plane $P_1H_2P_2O$. The light vectors at P_1' and P_2' are in phase and of equal amplitude. The disturbances they produce at O are OV_1 and OV_2, whose components interfere destructively along OX: the components along OY and OZ reinforce. The disturbances produced by the corresponding points in adjacent quadrants also cancel along OX and reinforce along OY and OZ. However, the components in the OY direction are in the opposite sense to those produced by P_1' and P_2', whereas along OZ all four components are in the same sense. Thus it is necessary to consider only the

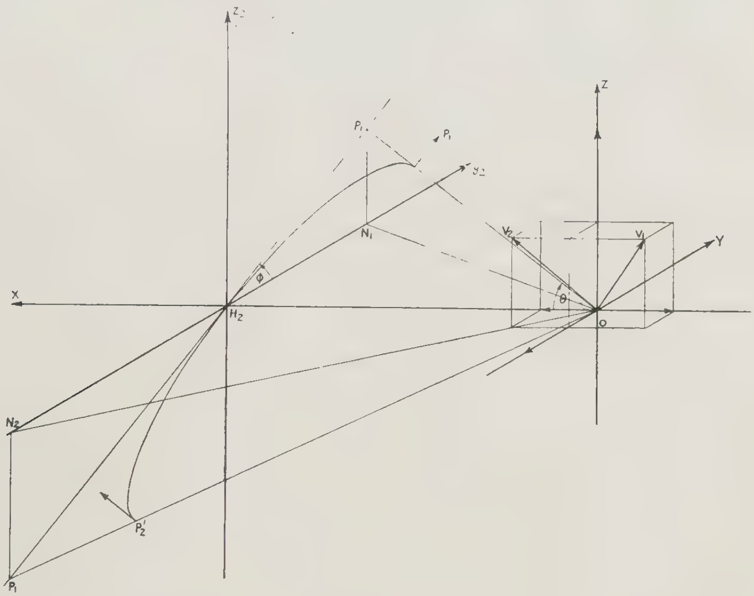


Figure 3.

components of the disturbances at O in the OZ direction, and the effective disturbance due to P_1' is thus equal to $OV_1 \cos ZOV_1$. From figure 3 we have

$$\sin \widehat{ZOV_1} = \sin\theta . \sin\phi,$$

θ being the angle P_1OH_2. Hence the effective disturbance can be written as

$$OV_1(1-\sin^2\theta . \sin^2\phi)^{\frac{1}{2}}, \qquad \ldots\ldots(4)$$

and the effective amplitude of S_3 is given as

$$\cos^{3/2}\theta . (1-\sin^2\theta . \sin^2\phi)^{\frac{1}{2}},$$

which is asymmetrical in ϕ. The phase relations of the disturbances at points the same distance just off-axis are thus applied to different distributions of amplitude, and the final light distribution is not independent of ϕ'.

The points P_1, P_2 in figure 2 have polar co-ordinates (ρ, ϕ), (ρ', ϕ') in the planes YOZ, $y_2H_2z_2$ respectively. P_1 lies on the sphere

$$X^2+Y^2+Z^2=R^2,$$

where $R=OP_1$. The area of S_2 is $\rho\, d\rho\, d\phi$, and therefore, by (3) above, that of S_3 is $\cos^{3/2}\theta\,.\,\rho\, d\rho\, d\phi$. The amplitude due to S_3 is thus $\cos^{3/2}\theta\,.\,\rho\, d\rho\, d\phi$, and (omitting constant factors) the disturbance OV_1 at the point P_2 is

$$OV_1 = \cos^{3/2}\theta\,.\,\sin 2\pi\left(\frac{t}{T}-\frac{d}{\lambda}\right)\rho\, d\rho\, d\phi,$$

in which $d=P_1P_2$. The component of OV_1 along OZ is, according to (4),

$$(OV_1)_z = \cos^{3/2}\theta\,.\,(1-\sin^2\theta\,.\,\sin^2\phi)^{\frac{1}{2}}\,.\,\sin 2\pi\left(\frac{t}{T}-\frac{d}{\lambda}\right)\rho\, d\rho\, d\phi. \qquad \ldots\ldots(5)$$

The co-ordinates of P_1, P_2 are $(x, \rho_1\cos\phi, \rho_1\sin\phi)$, $(0, \rho'\cos\phi', \rho'\sin\phi')$ with respect to the $O(XYZ)$ system. Thus

$$\begin{aligned} d^2 &= X^2+(\rho_1\cos\phi-\rho'\cos\phi')^2+(\rho_1\sin\phi-\rho'\sin\phi')^2 \\ &= R^2-2\rho_1\rho'\cos(\phi-\phi'), \end{aligned}$$

and to a near enough approximation, ρ' being small,

$$d=R-\rho\rho'/R\cos\theta\,.\,\cos(\phi-\phi'),$$

in which ρ_1 is replaced by $\rho\cos\theta$ (see figure 2 *a*).

If this value of d be substituted in (5) and the sine expanded in terms of sines and cosines of the constant and variable parts of the argument, it follows that the intensity at P_2 due to the element S_2 can be expressed as the squared modulus of the complex function dE, where

$$dE=\cos^{3/2}\theta\,.\,(1-\sin^2\theta\,.\,\sin^2\phi)^{\frac{1}{2}}\,e^{i\,.\,2\pi/\lambda\,.\,\rho\rho'/R\,.\,\cos\theta\,.\,\cos(\phi-\phi')}\rho\, d\rho\, d\phi.$$

The intensity I at P_2 due to the whole emergent wave is now given by

$$\sqrt{I}=\int_0^r \rho\cos^{3/2}\theta\,.\,d\rho\int_0^{2\pi}(1-\sin^2\theta\,.\,\sin^2\phi)^{\frac{1}{2}}\,.\,e^{ix\cos(\phi-\phi')}\,d\phi, \qquad \ldots\ldots(6)$$

r being the radius of the exit-pupil, and x being written for

$$2\pi/\lambda\,.\,\rho\rho'/R\,.\,\cos\theta.$$

Since $\rho=R\tan\theta$,

$$x=\frac{2\pi\rho'}{\lambda}\,.\,\sin\theta$$

and

$$\rho d\rho=R^2\,.\,\sin\theta\,.\,\cos^{-3}\theta\,.\,d\theta.$$

Dropping the constant factor R^2, substitution of this value of $\rho\, d\rho$ in (6) gives

$$\sqrt{I}=\int_0^\alpha \cos^{-3/2}\theta\,.\,\sin\theta d\theta\int_0^{2\pi}(1-\sin^2\theta\,.\,\sin^2\phi)^{\frac{1}{2}}\,.\,e^{ix\cos(\phi-\phi')}\,d\phi, \qquad \ldots\ldots(7)$$

where α is the limiting value of θ.

Let z be the value of x for $\theta=\alpha$, then

$$z=\frac{2\pi\rho'}{\lambda}\,.\,\sin\alpha \qquad \ldots\ldots(8)$$

and

$$\sin\theta=\frac{\sin\alpha}{z}\,.\,x.$$

Using this value for $\sin\theta$, (7) can be written

$$\sqrt{I}=\left(\frac{\sin\alpha}{z}\right)^2\int_0^z\left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\right)^{-5/4}.\,x\,dx\int_0^{2\pi}\left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\sin^2\phi\right)^{\frac{1}{2}}.\,e^{ix\cos(\phi-\phi')}d\phi. \qquad \ldots\ldots(9)$$

Expanding the square root by means of the binomial theorem, the integration with respect to ϕ becomes

$$\int_0^{2\pi}e^{ix\cos(\phi-\phi')}\,d\phi-\sum_{r=1}^{\infty}\frac{1}{2r}\left(\frac{1}{2}\cdot\frac{3}{4}\cdot\ldots\ldots\frac{2r-3}{2r-2}\right)\frac{\sin^{2r}\alpha}{z^{2r}}\,.\,x^{2r}\int_0^{2\pi}\sin^{2r}\phi\,.\,e^{ix\cos(\phi-\phi')}\,d\phi$$

$$=2\pi J_0(x)-\sum_{r-1}^{\infty}\frac{1}{2r}\left(\frac{1}{2}\cdot\frac{3}{4}\cdot\ldots\ldots\frac{2r-3}{2r-2}\right).\frac{\sin^{2r}\alpha}{z^{2r}}\,.\,x^{2r}\,I_{2r}, \qquad \ldots\ldots(10)$$

in which

$$I_{2r}=\int_0^{2\pi}\sin^{2r}.\,\phi\,.\,e^{ix\cos(\phi-\phi')}\,d\,\phi.$$

To discuss the case $\phi'=0$, consider Bessel's second integral,* which can be written

$$J_r(x)=\frac{1}{\sqrt{\pi}\,\Gamma(r+\frac{1}{2})}\left(\frac{x}{2}\right)^r\int_0^{\pi}e^{\pm ix\cos\phi}.\,\sin^{2r}\phi\,d\phi.$$

Expanding the gamma function and changing the limits of integration, this becomes

$$\int_0^{2\pi}\sin^{2r}\phi\,.\,e^{\pm ix\cos\phi}\,d\phi=(2r-1)(2r-3)\ldots\ldots3\,.\,1\,.\,2\pi\,.\,x^{-r}J_r(x).$$

Thus (10) can be written

$$2\pi J_0(x)-2\pi\sum_{r=1}^{\infty}c_r\,.\,\frac{\sin^{2r}\alpha}{z^{2r}}\,.\,x^rJ_r(x), \qquad \ldots\ldots(11)$$

the coefficient being

$$c_r=\frac{1}{2r}\left(\frac{1}{2}\cdot\frac{3}{4}\cdot\ldots\ldots\frac{2r-3}{2r-2}\right)(2r-1)(2r-3)\ldots\ldots3\,.\,1, \qquad c_1=\tfrac{1}{2}.$$

Let the function $f(x^2)$ be defined as

$$f(x^2)=\cos^{5/2}\alpha\left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\right)^{-5/4};$$

then, apart from the factor $2\pi\cos^{-5/2}\alpha\,.\,\sin^2\alpha$, equation (9) is

$$\sqrt{I}=\frac{1}{z^2}\int_0^z f(x^2)\,.\,xJ_0(x)-\sum_{r=1}^{\infty}c_r\,.\,\frac{\sin^{2r}\alpha}{z^{2r+2}}\int_0^z f(x^2)\,.\,x^{r+1}J_r(x)dx,$$

which becomes, on integration,†

$$\frac{1}{z^2}\left\{f(z^2)\,.\,zJ_1(z)+(-2)f^{(1)}(z^2)\,.\,z^2J_2(z)+\ldots\ldots+(-2)^{k-1}f^{(k-1)}(z^2)\,.\,z^kJ_k(z)+\ldots\ldots\right\}$$

$$-\sum_{r=1}^{\infty}c_r\,.\,\frac{\sin^{2r}\alpha}{z^{2r+2}}\left\{f(z^2)\,.\,z^{r+1}J_{r+1}(z)+(-2)f^{(1)}(z^2)\,.\,z^{r+2}J_{r+2}(z)+\ldots\ldots+(-2)^{k-1}f^{(k-1)}(z^2)z^{r+k}J_{r+k}(z)+\ldots\ldots\right\}.$$

* See Gray and Mathews, *Bessel Functions*, p. 46.

† See G. C. Steward, *Philos. Trans.* A, 1926 "Aberrational Diffraction Effects" (Appendix).

Introduce now a function u_s, defined by

$$u_s = z^{2s} . f^{(s)}(z^2), \qquad u_0 = f(z^2), \qquad \ldots\ldots(12)$$

then

$$\sqrt{I} = \left\{ u_0 . \frac{J_1(z)}{z} + (-2)u_1 . \frac{J_2(z)}{z^2} + (-2)^2 u_2 . \frac{J_3(z)}{z^3} + \ldots\ldots \right\}$$

$$- \sum_{z=1}^{\infty} c_r \sin^{2r}\alpha \left\{ u_0 . \frac{J_{r+1}(z)}{z^{r+1}} + (-2)u_1 . \frac{J_{r+2}(z)}{z^{r+2}} + (-2)^2 u_2 . \frac{J_{r+3}(z)}{z^{r+3}} + \ldots\ldots \right\},$$

or collecting like terms,

$$\sqrt{I} = u_0 \frac{J_1(z)}{z} - (2u_1 + c_1 u_0 \sin^2\alpha) . \frac{J_2(z)}{z^2} + (4u_2 + 2c_1 u_1 \sin^2\alpha - c_2 u_0 \sin^4\alpha) . \frac{J_3(z)}{z^3}$$

$$- (8u_3 + 4c_1 u_2 \sin^2\alpha - 2c_2 u_1 \sin^4\alpha + c_3 u_0 \sin^6\alpha) . \frac{J_4(z)}{z^4} + \ldots\ldots \qquad \ldots\ldots(13)$$

The values of u_0, u_1 ...... etc., can be found from the equations

$$f(t) = \cos^{5/2}\alpha \left(1 - \frac{\sin^2\alpha}{z^2} . t\right)^{-5/4},$$

$$f^{(1)}(t) = \cos^{5/2}\alpha \frac{5}{4} \left(1 - \frac{\sin^2\alpha}{z^2} . t\right)^{-9/4} . \frac{\sin^2\alpha}{z^2},$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$f^{(k)}(t) = \cos^{5/2}\alpha . \frac{5}{4} . \frac{9}{4} \ldots\ldots \frac{1+4k}{4} \left(1 - \frac{\sin^2\alpha}{z^2} . t\right)^{-\frac{8+4k}{4}} \left(\frac{\sin^2\alpha}{z^2}\right)^k,$$

for it follows from the definition (12) that

$$\left.\begin{aligned} u_0 &= 1, \\ u_1 &= \frac{5}{4} . \tan^2\alpha, \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ u_k &= \frac{5}{4} . \frac{9}{4} \ldots\ldots \frac{1+4k}{4} . \tan^{2k}\alpha, \end{aligned}\right\} \qquad \ldots\ldots(14)$$

and the coefficients in (13) are independent of z. Equations (13) and (14) provide a formal solution of the problem valid for systems of high relative aperture. Equation (13) is of the form

$$\sqrt{I} = \frac{J_1(z)}{z} + a_2 \frac{J_2(z)}{z^2} + a_3 \frac{J_3(z)}{z^3} + a_4 \frac{J_4(z)}{z^4} + \ldots\ldots, \qquad \ldots\ldots(15)$$

the constants a_2 ...... being determined solely by the angular semi-aperture α. The values of these constants for every 10° up to 40° are given in table 1. When

Table 1. Constants for Gaussian system $\phi' = 0$

α	$\sin\alpha$	a_2	a_3	a_4
10°	·1736	−0·0928	0·0119	− 0·0024
20°	·3420	−0·3897	0·2117	− 0·2369
30°	·5000	−0·9583	1·3308	− 2·8596
40°	·6428	−1·9668	5·8774	−26·628

$\sin\alpha > 0{\cdot}65$, the formulae become increasingly inconvenient for accurate numerical calculation, and the values of the constants for $\alpha > 40°$ have, therefore, not been calculated.

When $\phi' = \dfrac{\pi}{2}$ it is simpler to proceed as follows. The integration with respect to ϕ in (9) can be written

$$\int_0^{2\pi}\left\{\left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\right)+\frac{\sin^2\alpha}{z^2}\,.\,x^2\cos^2\phi\right\}^{\frac{1}{2}} e^{ix\cos(\phi-\phi')}\,d\phi.$$

Put $\phi' = \pi/2$ and expand the square root to give

$$\left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\right)^{\frac{1}{2}}\int_0^{2\pi} e^{ix\sin\phi}\,d\phi + \sum_{r=1}^{\infty}\frac{\frac{1}{2}(\frac{1}{2}-1)\,.\,.\,.\,.\,.\,(\frac{1}{2}-\overline{r-1})}{r!}\left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\right)^{\frac{1}{2}-r}$$
$$.\,\frac{\sin^{2r}\alpha}{z^{2r}}\,.\,x^{2r}\int_0^{2\pi}\cos^{2r}\phi\,.\,e^{ix\sin\phi}\,d\phi,$$

which is equal to

$$\left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\right)^{\frac{1}{2}} 2\pi J_0(x) + \sum_{r=1}^{\infty}\frac{\frac{1}{2}(\frac{1}{2}-1)\,.\,.\,.\,.\,.\,(\frac{1}{2}-\overline{r-1})}{r!}\left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\right)^{\frac{1}{2}-r}$$
$$.\,\frac{\sin^{2r}\alpha}{z^{2r}}\,.\,x^{2r}\,.\,1\,.\,3\,.\,.\,.\,.\,.\,(2r-1)x^{-r}J_r(x).$$

Thus the expression (9) can be written

$$\sqrt{I} = 2\pi\left(\frac{\sin\alpha}{z}\right)^2\int_0^z\left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\right)^{-\frac{3}{4}} xJ_0(x)\,dx$$
$$+2\pi\sum_{r=1}^{\infty}\frac{\frac{1}{2}(\frac{1}{2}-1)\,.\,.\,.\,.\,.\,(\frac{1}{2}-\overline{r-1})}{r!}\,1\,.\,3\,.\,.\,.\,.\,.\,(2r-1)\left(\frac{\sin\alpha}{\alpha}\right)^{2r+2}$$
$$.\int_0^z\left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\right)^{-\frac{3}{4}-r} x^{r+1}J_r(x)\,dx.$$

Let the function $f_r(x^2)$ be defined by

$$f_r(x^2) = \cos^{3/2}\alpha\left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\right)^{-\frac{3}{4}-r},$$

so that, apart from the factor $2\pi\sin^2\alpha\,.\,\cos^{-3/2}\alpha$, the intensity is given by

$$\sqrt{I} = \frac{1}{z^2}\int_0^z f_0(x^2)xJ_0(x)\,dx + \sum_{r=1}^{\infty}\frac{\frac{1}{2}(\frac{1}{2}-1)\,.\,.\,.\,.\,.\,(\frac{1}{2}-\overline{r-1})}{r!}\,.\,1\,.\,3\,.\,.\,.\,.\,.\,(2r-1)$$
$$.\,\sin^{2r}\alpha\,\frac{1}{x^{2r+2}}\int_0^z f_r(x^2)\,.\,x^{r+1}J_r(x)\,dx$$
$$=\frac{1}{z^2}\left\{f_0(z^2)\,.\,zJ_1(z)+(-2)f_0^{(1)}(z^2)\,.\,z^2J_2(z)+(-2)^2f_0^{(2)}(z^2)\,.\,z^3J_3(z)+\,.\,.\,.\,.\,.\,.\right\}$$
$$+\sum_{r=1}^{\infty}\frac{\frac{1}{2}(\frac{1}{2}-1)\,.\,.\,.\,.\,.\,(\frac{1}{2}-\overline{r-1})}{r!}\,.\,1\,.\,3\,.\,.\,.\,.\,.\,(2r-1)\sin^{2r}\alpha\,\frac{1}{z^{2r+2}}$$
$$.\left\{f_r(z^2)\,.\,z^{r+1}J_{r+1}(z)+(-2)f_r^{(1)}(z^2)z^{r+2}J_{r+2}(z)+\,.\,.\,.\,.\,.\,.\right\}. \qquad \ldots\ldots(16)$$

Introduce functions $(u_0)_s$, $(u_r)_s$, defined by the relations

$$\left.\begin{aligned}(u_0)_s &= z^{2s}f_0^{(s)}(z^2), & (u_0)_0 &= f_0(z^2),\\ (u_r)_s &= z^{2s}f_r^{(s)}(z^2)\,.\,\cos^{2r}\alpha, & (u_r)_0 &= f_r(z^2)\,.\,\cos^{2r}\alpha.\end{aligned}\right\}\ \ldots\ldots(17)$$

Then the intensity as given by (16) becomes

$$\sqrt{I}=\left\{(u_0)_0\frac{J_1(z)}{z}+(-2)(u_0)_1\frac{J_2(z)}{z^2}+(-2)^2(u_0)_2\,.\frac{J_3(z)}{z^3}+\ldots\ldots\right\}$$

$$+\sum_{z=1}^{\infty}\frac{\frac{1}{2}(\frac{1}{2}-1)\ldots\ldots(\frac{1}{2}-\overline{r-1})}{r!}\,.1\,.3\ldots\ldots(2r-1)\tan^{2r}\alpha$$

$$.\left\{(u_r)_0\frac{J_{r+1}(z)}{z^{r+1}}+(-2)(u_r)_1\frac{J_{r+2}(z)}{z^{r+2}}+\ldots\ldots\right\},$$

or, collecting like terms,

$$\sqrt{I}=(u_0)_0\frac{J_1(z)}{z}-\left[2(u_0)_1-\frac{1}{2}\tan^2\alpha(u_1)_0\right].\frac{J_2(z)}{z^2}$$

$$+\left[4(u_0)_2-\tan^2\alpha\,.\,(u_1)_1-\frac{3}{8}\tan^4\alpha\,.\,(u_2)_0\right].\frac{J_3(z)}{z^3}$$

$$-\left[8(u_0)_3-2\tan^2\alpha(u_1)_2-\frac{3}{4}\tan^4\alpha(u_2)_1+\frac{15}{16}\tan^6\alpha(u_3)_0\right]\frac{J_4(z)}{z^4}+\ldots\ldots \qquad \ldots\ldots(18)$$

The values of $(u_0)_0$, $(u_0)_1\ldots\ldots$, $(u_r)_0$, $(u_r)_1$, ... etc., can be found from the equations (17). They are

$$(u_0)_0=1, \qquad (u_r)_0=1,$$

$$(u_0)_1=\tfrac{3}{4}\,.\tan^2\alpha, \qquad (u_r)_1=(\tfrac{3}{4}+r)\tan^2\alpha,$$

$$\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot \qquad \cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot$$

$$(u_0)_k=\frac{3}{4}\cdot\frac{7}{4}\cdot\ldots\ldots\frac{4k-1}{4}\tan^{2k}\alpha, \qquad (u_r)_k=(\tfrac{3}{4}+r)(\tfrac{3}{4}+r+1)\ldots\ldots$$

$$(\tfrac{3}{4}+r+k-1)\tan^{2k}\alpha, \qquad \ldots\ldots(19)$$

which are independent of z. Using these values in (18), this latter becomes

$$\sqrt{I}=\frac{J_1(z)}{z}-\tan^2\alpha\,.\frac{J_2(z)}{z^2}+\frac{25}{8}\tan^4\alpha\,.\frac{J_3(z)}{z^3}-\frac{203}{8}\tan^5\alpha\,.\frac{J_4(z)}{z^4}+\ldots\ldots \qquad \ldots\ldots(20)$$

which is of the same form as (15). The constants $a_2\ldots\ldots$ are again determined solely by the angular semi-aperture α: their values for every 10° up to 40° are given in table 2.

Table 2. Constants for Gaussian system $\phi'=\pi/2$

α	$\sin\alpha$	a_2	a_3	a_4
10°	·1736	−·0311	+0·0030	−0·0008
20°	·3420	−·1325	+0·0548	−0·0782
30°	·5000	−·3333	+0·3472	−0·9399
40°	·6428	−·7041	+1·549	−8·858

If the approximation be made that $\cos\alpha=1$, then α^2, $\sin^2\alpha$ and $\tan^2\alpha$ are negligible, and the equations (13) and (20) both reduce to

$$\sqrt{I}=\frac{J_1(z)}{z}, \qquad \ldots\ldots(21)$$

which is the Airy disc formula (1). Figures 4 and 5 show the intensity curves according to (13) and (20) for $\alpha=40°$ in the cases $\phi'=0$ and $\phi'=\pi/2$, together with the curves of (21). These and tables 1 and 2 show that the Airy disc formula is accurate to within one or two per cent for $\alpha<30°$, providing the argument of the function (21) is

$$z=\frac{2\pi}{\lambda}\sin\alpha\,.\,\rho'. \qquad \ldots\ldots(22)$$

Further, for $\alpha>30°$ the resolution of non-coherent star images is slightly greater than that given by (21) and (22), but for $\alpha<30°$ the Rayleigh formula, written

$$d=0.61\,\frac{\lambda}{\sin\alpha}, \qquad \ldots\ldots(23)$$

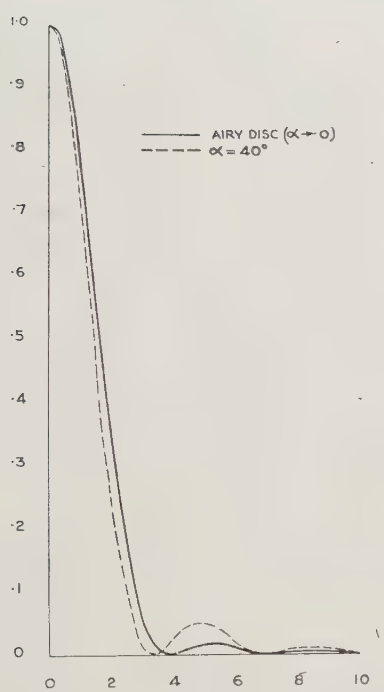


Figure 4. Distribution for a Gaussian system in a meridian perpendicular to the direction of the incident light vector ($\phi'=0$).

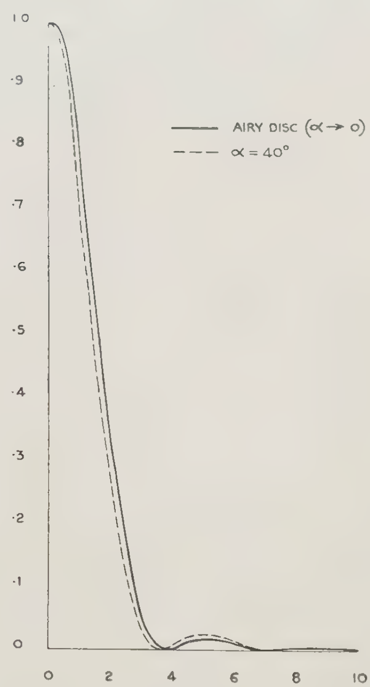


Figure 5. Distribution for a Gaussian system in a meridian parallel with the direction of the incident light vector ($\phi'=\pi/2$).

is remarkably accurate. Thus, to be applicable to systems of higher aperture, the forms (21), (22) and (23) must be employed, using $\sin\alpha$ in place of the ratio of the radius of the exit pupil to the focal length of the system, which in extreme cases leads to an absurdly high degree of resolution.

§ 2. SYSTEMS OBEYING THE SINE CONDITION

In a Gaussian system the unit surfaces are planes through the paraxial unit points, and in such a system the energy passing through an annulus $(\rho, \rho+d\rho)$ on the object side will emerge through an annulus between these radii on a plane,

through the unit point on the image side. This is not the case in a system which fulfils the sine condition; for the appropriate unit surface is then not a plane, but the emergent wave-front itself, assuming an object point at infinity. The incident wave being plane, independently of the form of the unit surface on the incident side, the energy entering through the annulus $(\rho, \rho+d\rho)$ will all pass through the annulus on the emergent wave-surface which lies between the circles formed by intersection of this latter with two cylinders concentric with the axis and of radii ρ and $\rho+d\rho$. Thus the amplitude over the element S_2 of the emergent wave-front (figure 6) is equal to

$$\left\{\frac{\text{area } S_1}{\text{area } S_2}\right\}^{\frac{1}{2}} = \cos^{\frac{1}{2}}\theta.$$

Hence in a system obeying the sine condition, the amplitude over the emergent wave-front decreases with increasing values of θ.

Equation (6) will now take the form

$$\sqrt{I} = \int_0^r \rho\cos^2\theta\, d\rho \int_0^{2\pi} (1-\sin^2\theta\,.\,\sin^2\phi)^{\frac{1}{2}}\, e^{ix\cos(\phi-\phi')}\, d\phi,$$

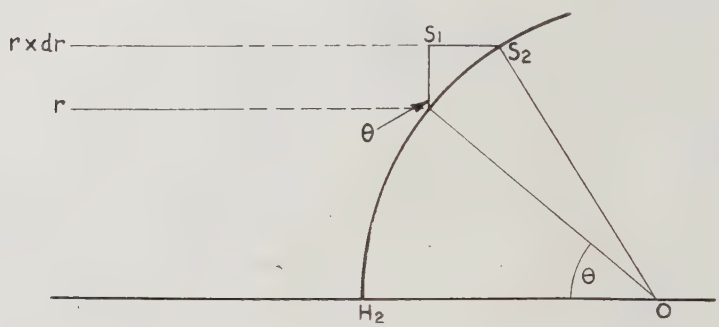


Figure 6

which reduces to

$$\sqrt{I} = \left(\frac{\sin\alpha}{z}\right)^2 \int_0^z \left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\right)^{-\frac{1}{4}} x\,dx \int_0^{2\pi} (1-\sin^2\theta\,.\,\sin^2\phi)^{\frac{1}{2}}\, e^{ix\cos(\phi-\phi')}\, d\phi, \qquad \text{......(24)}$$

in which the integration with respect to ϕ is identical with that treated above.

In the case $\phi'=0$, put $g(x^2)$ in place of $f(x^2)$, where

$$g(x^2) = \cos^{\frac{1}{2}}\alpha\left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\right)^{-\frac{1}{4}}$$

Then the intensity can be written exactly as in (13), if

$$u_s = z^{2s}g^{(s)}(z^2), \qquad u_0 = g(z^2).$$

The values of u_0, u_1......, etc., are given by

$$\left.\begin{aligned} u_0 &= 1, \\ u_1 &= \tfrac{1}{4}\,.\,\tan^2\alpha, \\ &\cdots\cdots \\ u_k &= \frac{1}{4}\cdot\frac{5}{4}\cdots\cdots\frac{4k-3}{4}\,.\,\tan^{2k}\alpha. \end{aligned}\right\} \qquad \text{......(25)}$$

Thus the intensity is again given by an equation of the form (15): the constants for this case are given in table 3.

Table 3. Constants for sine-condition system $\phi'=0$

a	$\sin a$	a_2	a_3	a_4
10°	·1736	−0·0306	+0·0010	− 0·0002
20°	·3420	−0·1247	+0·0177	− 0·0201
30°	·5000	−0·2916	+0·1363	− 0·2364
40°	·6428	−0·5586	+0·8284	− 2·1032
50°	·7660	−1·0035	+2·6001	−14·140

When $\phi'=\pi/2$, put $g_r(x^2)$ in place of $f_r(x^2)$, where

$$g_r(x^2) = \cos^{-\frac{1}{2}}\alpha\left(1-\frac{\sin^2\alpha}{z^2}\,.\,x^2\right)^{\frac{1}{2}-r}.$$

Then the intensity can be written exactly as in (18), if

$$(u_0)_s = z^{2s} g_0^{(s)}(z^2), \qquad (u_0)_0 = g_0(z^2),$$
$$(u_r)_s = z^{2s} g_r^{(s)}(z^2)\,.\,\cos^{2r}\alpha, \qquad (u_r)_0 = g_r(z^2)\,.\,\cos^{2r}\alpha.$$

The values of $(u_0)_0$, $(u_0)_1, \ldots\ldots, (u_r)_0, (u_r)_1, \ldots\ldots$, etc., are

$$(u_0)_0 = 1, \qquad (u_r)_0 = 1,$$
$$(u_0)_1 = -\tfrac{1}{4}\tan^2\alpha, \qquad (u_r)_1 = +(r-\tfrac{1}{4})\tan^2\alpha.$$
$$\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot \qquad \cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot$$
$$(u_0)_k = -\frac{1}{4}\cdot\frac{3}{4}\cdot\ldots\ldots\frac{4k-5}{4}\,.\,\tan^{2k}\alpha, \quad (u_r)_k = +(r-\tfrac{1}{4})(r+1-\tfrac{1}{4})\ldots\ldots(r+\overline{k-1}-\tfrac{1}{4})\tan^{2k}\alpha. \qquad \ldots\ldots(26)$$

Thus equation (20) becomes

$$\sqrt{I} = \frac{J_1(z)}{z} + \tan^2\alpha\,.\,\frac{J_2(z)}{z^2} - \frac{5}{6}\tan^4\alpha\,.\,\frac{J_3(z)}{z^3} + 3\tan^6\alpha\,.\,\frac{J_4(z)}{z^4} - \ldots\ldots, \qquad \ldots\ldots(27)$$

which is again of the form (15): the constants for this case are given in table 4.

Table 4. Constants for sine-condition system $\phi'=\pi/2$

a	$\sin a$	a_2	a_3	a_4
10°	·1736	+0·0311	−0·0008	+0·0001
20°	·3420	+0·1325	−0·0146	+0·0094
30°	·5000	+0·3333	−0·0926	+0·1111
40°	·6428	+0·7041	−0·4131	+1·0471
50°	·7660	+1·4203	−1·6808	+7·0948

Figures 8 and 9 show the intensity curves for a sine-condition system in the two cases $\phi'=0$ and $\phi'=\frac{\pi}{2}$. For $\alpha<30°$, the conclusions expressed in (21), (22) and (23) are again valid.

The curves of figures 7 and 8 have bearing on the resolving power of microscopes. A variation of the resolution for two fixed points with rotation of the direction of polarization is suggested by them.

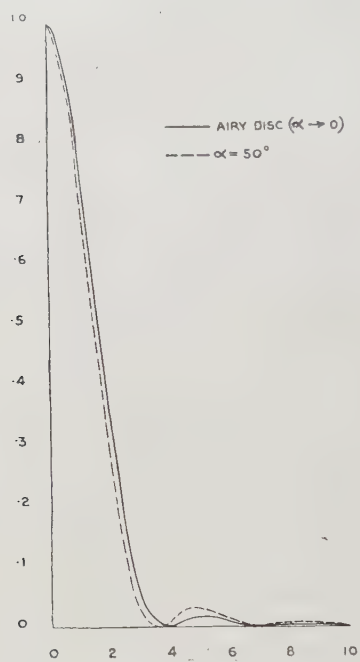


Figure 7. Distribution for a sine-condition system in a meridian perpendicular to the direction of the incident light vector ($\phi'=0$).

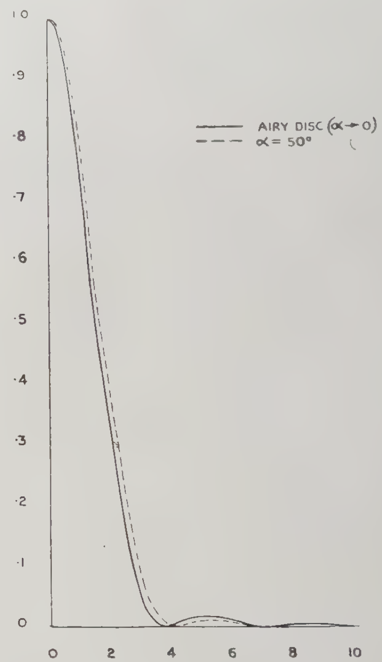


Figure 8. Distribution for a sine-condition system in a meridian parallel with the direction of the incident light vector ($\phi'=\pi/2$).

It should be noted that the effects in various planes of focus and in the presence of spherical aberration of any order can be investigated by the method, for only the forms of the u-coefficients are modified, and equation (18) is then directly applicable to imperfectly corrected microscope-objectives under different conditions of focus.

Plate 1. Newton at about the time of writing the *Principia* ; engraved by Thomas Oldham Barlow, after Kneller.

[*By permission of Trinity College, Cambridge, from the engraving in their possession.*]

Plate 2. Newton at the age of fifty-nine years, by Kneller.

[*By permission of the Trustees of the National Portrait Gallery.*]

To face page 129

NEWTON

By E. N. da C. ANDRADE, F.R.S.,
London

Newton Tercentenary Lecture, delivered in London 4 *December and in Cambridge* 9 *December* 1942

Just three hundred years ago Newton was born at Woolsthorpe, a Lincolnshire hamlet. I will not trouble you with the history of his childhood; the only thing about it that seems significant to one unskilled to find the whole future greatness of a man foreshadowed in unauthentic anecdotes of his boyhood is that he took a great delight in making mechanical toys and devices of all kinds, such as the model windmill of which the pious Stukeley tells us. Every good experimenter, I think, has shown this youthful liking for making things with his own hands. Robert Hooke as a boy spent his time making little mechanical toys: Clerk Maxwell's childish inventiveness in doing and making is recorded in his *Life*: while among the toys and models that the child Heinrich Hertz made was a mill that worked. A strong bent for making things himself continued to be characteristic of Newton as long as he was active in science. As a young man in his early twenties at Cambridge he spent much time figuring and grinding lenses; the second reflecting telescope that he made with his own hands is still in the possession of the Royal Society; and Humphrey Newton, his assistant and amanuensis from 1685 to 1690, tells us that, during this period, the brick furnaces, which he used for his alchemical experiments "*pro re nata*"—as to the manner born—"he made and altered himself without troubling a bricklayer." He said to Conduitt, who asked him about the making of his telescope: "If I had stayed for other people to make my tools and things I had never made anything of it". He also, as a young man, took an interest in drawing and painting as a craft, but there is nothing recorded to indicate that he took any interest in art as such—in fact, later in life he referred to Lord Pembroke's splendid collection of Greek statuary as "My Lord's old fashion'd babys"* and to Pembroke himself as "a lover of stone dolls". In a musical age he never seems to have taken much interest in music, although, speaking of a period towards the end of his life, Stukeley says he was fond of music, which he follows by an anecdote of how Newton said: "I went to the last opera. The first act gave me the greatest pleasure. The second quite tired me: at the third I ran away." As we do not know what the opera was, we cannot judge of his taste.

Going to Cambridge at the age of eighteen, he spent his first two years learning the ordinary mathematics of his time—Euclid and trigonometry—and then, in 1663, came under the influence of Barrow, one of the best mathematicians of his age, a man keenly interested in optics, who in his *Lectiones Opticae*, which appeared in 1669, acknowledged Newton's help in the Preface.

* Stukeley, p. 66. In Newton's time the word "baby" was commonly used for "doll".

He read Kepler's *Dioptrice*, which was probably the best textbook of geometrical optics available; Descartes' *Geometria*, which described the new analytical geometry; Schooten's *Exercitationum Mathematicarum Libri V*, a mathematical miscellany; and Wallis's works, including the *Arithmetica Infinitorum*, which closely foreshadowed the calculus. Thus, by the time the plague drove him

Newton's Birth-Place at Woolsthorpe.

These premises, the paternal residence of Newton, the house in which that great man was born, and in which he was reared and educated by his widowed mother, stands in a small romantic valley, about a mile west of the North Road, and five miles south of Grantham. At the school of this latter town he was educated, and used to walk daily to it, like other country boys, with his satchel at his back, or, after he was boarding in the town, on Monday mornings. The stone-work of the house is still scratched with several rude dials of his early formation, on some of which also remain his own clumsy gnomons. His bed-room too is shown, as well as the corner separated by a partition constructed by himself, to serve as a study. It is the right-hand window on the first story; and he was born in the room beneath, on the ground-floor. The apple tree, whence he drew his hypothesis, *about gravitation, is still in existence and bearing, but bent to the ground and broken, as represented in the right-hand corner; and the pear-tree under which he used to sit, and which the residents assert was the subject of the above observation, is still standing and flourishing, both within fifty yards of the door in the dwelling.*

Figure 1. From an old leaflet figuring and describing Newton's birthplace.

from Cambridge in 1665, he was already master of the most recent advances in mathematics and optics and had, he himself tells us, "found the method of Infinite Series", which from the context implies the principle of integral calculus, since with it he computed the area of the hyperbola to fifty-two figures, one of many instances of a liking for playing with figures for its own sake.

The two years 1665 and 1666, most of which Newton spent at Woolsthorpe, whither he had been driven from Cambridge by the plague, were the great

Plate 3.

Portrait of Newton "when Bachelor of Arts in Trinity College, Cambridge". The engraving here represented is from a painting by Lely, at one time in the possession of Lord Cremorne. The history of the painting appears to be unknown.

Figure 2. The reflecting telescope, made by Newton with his own hands, now in the possession of the Royal Society.

To face page 131

springtime of his genius. I cannot refrain from quoting the famous memorandum in the Portsmouth Collection, written about 1714, where, after describing his early work on gravity as an interplanetary force, he adds: "All this was in the two plague years of 1665 and 1666, for in those days I was in the prime of my age for invention, and minded mathematics and philosophy more than at any time since." When the young man of twenty-four returned to Cambridge early in 1667 he had already laid the general foundations of the differential and integral calculus, he had a pretty clear idea that terrestrial gravity, acting according to an inverse square law, controlled the moon's motion, he had carried out fundamental experiments on the refraction of light with the prism and he had begun to construct a reflecting telescope. He had, however, published nothing, and seems not to have had publication in mind. In fact, although Newton had gone through the manuscript of Barrow's *Lectiones Opticae*, published in 1669, making certain corrections and additions, and had read the proofs, there is no mention in the book of his work on the reflecting telescope or the prism.*

In 1667 he was elected a Minor Fellow and, in 1668, a Major Fellow of Trinity. It may be mentioned that there were several vacancies among the Fellows in the former year, two of them due to Fellows having fallen down staircases, while one of the Senior Fellows had been put out from College on account of insanity. Whether the rarity of such an event in these days is due to a greater leniency or to a stricter sobriety and sanity among the Fellows we will not inquire. In 1669 Barrow retired from his Chair, and Newton, on his nomination, succeeded him as Lucasian Professor of Mathematics, at the age of twenty-six. In the years immediately following he gave a series of lectures on optics which included his chief discoveries in the matter of the prismatic decomposition of white light. A manuscript copy of the lectures was duly deposited in the University archives, but it was not until after his death that they were published. At the time they apparently attracted no attention, nor did the lecturer take any steps to get the work known.

Newton first came before the world of science, as represented by the Royal Society, in consequence of the reflecting telescope which he had made—and made because his experiments with prisms and with lenses, which he ground himself, had convinced him that chromatic aberration was the main cause of the imperfection of images formed by refraction, and that it was inherent in lenses. The Royal Society heard of this telescope and urgently wished to see it. For the Society he made a second instrument (figure 2), which is still in its possession as one of its most precious belongings. It aroused great interest, which was the occasion of Newton writing to the Society that "had not the communication of it been desired, I might have let it still remain private as it hath already done some years", the first expression of the typical Newtonian attitude towards the publication of his discoveries. Remember that at the time he had the fundamentals of the calculus in writing, unpublished. However, he expresses his gratification at his forthcoming election into the Society as a Fellow.

* The actual title of Barrow's book, usually referred to as *Lectiones Opticae*, is *Lectiones XVIII, in quibus Opticorum Phaenomenωn genuinae Rationes investigantur ac exponuntur.*

The consequence of the excitement produced by the telescope was that Newton offered to lay before the Society "an account of a philosophical discovery, which induced me to make the said telescope, and which I doubt not but will prove much more grateful than the communication of the instrument, being in my judgment the oddest if not the most considerable detection which hath hitherto been made in the operations of nature"—a very exceptional enthusiasm. Accordingly he sent a communication which was published in the *Philosophical Transactions* for February 19, 1671/72. It is headed : "*A letter of Mr.* Isaac Newton, *Mathematick Professor in the University of Cambridge; containing his New Theory about* Light *and* Colors; *Where* Light *is declared to be not Similar or Homogeneal, but consisting of difform rays, some of which are more refrangible than others: And* Colors *are affirmed to be not Qualifications of Light, deriv'd from Refractions of natural Bodies (as 'tis generally believed;) but Original and Connate properties, which in divers rays are divers: Where several Observations and Experiments are alledged to prove the said Theory.*" It may be remarked that the paper opens by saying that it was at the beginning of 1666, when he was busy grinding optic glasses of other figures than spherical, that he carried out his first experiment with the prism. The "solemn thanks of the meeting were voted to its author for his very ingenious discourse". Newton was launched on his public career, and was soon to know the penalties which attended its pleasures. Incidentally it may be remarked that about this time appeared the edition of the *Geographia* of Varenius which was edited by Newton. Apparently the book was used as a textbook for some lectures which he delivered.

The paper on the prism aroused great interest but also much criticism, in particular from Ignatius Pardies, Franciscus Linus and Hooke. Pardies was soon convinced by Newton; Linus, who died in 1675, raised various objections, mostly trivial: he seems to have spent much of his time bringing absurd objections against discoveries of prime importance, as witness his controversies with Boyle about the effects of a vacuum. After him Gascoigne and Lucas offered other criticisms, but it was Hooke who made the most trouble, as was to be expected of so acute a mind. He confirmed and admitted the constant refraction of a beam of a given isolated spectral colour, but maintained his view that all colours could be reduced to a mixture of red and blue, in support of which he produced certain arguments, shrewd, indeed, but which Newton had little trouble in demolishing. Huygens also misunderstood Newton's point of view, which was, in Newton's words ". . . the Theory, which I propounded, was evinced by me, not by inferring 'tis thus because not otherwise, that is, not by deducing it only from a confutation of contrary suppositions, but by deriving it from Experiments concluding positively and directly". The result of all these misunderstandings both of his experiments and of the conclusions that could safely be drawn from them was to irritate Newton in the highest degree: he wrote to Leibnitz in 1675: "I was so persecuted with discussions arising from the publication of my theory of light, that I blamed my own prudence for parting with so substantial a blessing as my quiet to run after a shadow." When Newton reached a conclusion he was so certain and it was all so clear to him that he found it difficult, I think, to believe that anyone of ability could honestly differ from him: he was apt to put down all questions

and doubts to deliberate persecution. He did not so much mind, I think, having his theories attacked, but he could not bear to have his experimental results questioned, and he could not understand what he properly regarded as futile arguments about what ought to happen. He was morbidly sensitive, too, about any suggestion that he had taken from others without proper acknowledgement.

Nevertheless, at the end of 1675 he sent to the Royal Society a further paper describing a theory "containing, partly, an Hypothesis to explain the properties of light discoursed of by him in his former papers, partly the principal phenomena

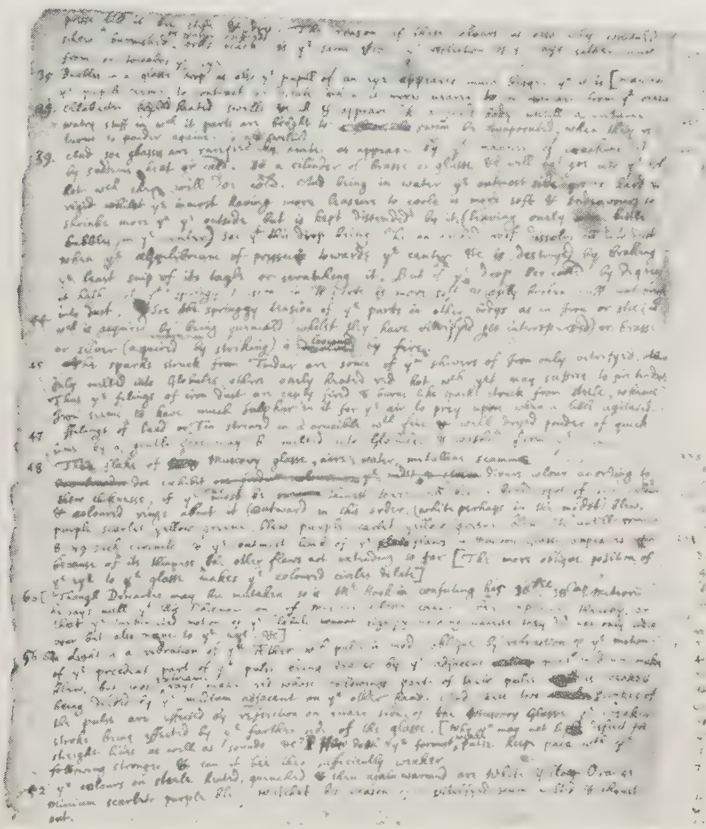


Figure 3. A sheet of Newton's notes on Hooke's *Micrographia*.
(*Reproduced by permission of the Syndics of the Cambridge University Library.*)

of the various colours exhibited by thin plates and bubbles, esteemed to be of a more difficult consideration, yet to depend also on the said properties of light". This paper describes the theory of fits of easy reflection and refraction, about which I will say a few words later, and discusses the colours of thin plates, including Newton's rings. In connection with this theory Newton has to speak of the properties of an aether which he postulates, and incidentally describes an electrical experiment, on the dancing to and fro of bits of paper and such like under a rubbed glass sheet. Hooke claimed that much of what Newton had done was in his *Micrographia*, published in 1665, and Newton admitted that he had made use of Hooke's observations on thin plates, but added, justly enough,

that he had carried the matter much further. That Newton had studied certain parts of the *Micrographia* very carefully is clear from seven sheets of holograph notes preserved in the Portsmouth Collection. One sheet is reproduced in figure 3.*

On this occasion Hooke wrote Newton an admirable letter, including the passage: "Your design and mine are, I suppose, both at the same thing, which is the discovery of truth, and I suppose we can both endure to hear objections, so they come not in a manner of open hostility, and have minds equally inclined to yield to the plainest deductions of reason from experiment." Newton answered generously enough. If a genial spirit like Halley had been at hand to sweeten the relations between the two difficult and touchy men further troubles might have been avoided; instead there was Oldenburg, ever anxious to do what he could against Hooke. Newton seldom did Hooke justice or acknowledged publicly what he did in fact owe him, and Hooke was mortified to find a younger man take what he had done and, with such apparent ease, go far beyond what, he probably felt in secret, he could do himself. At any rate, bad feeling soon arose again, with the result that Newton did not publish his *Opticks* until 1704, the year after Hooke died, and in it never mentioned Hooke, either in connection with thin plates or elsewhere. A letter to Boyle early in 1679 shows that he was still considering an all-pervading aether, in terms of which he explained diffraction and many chemical facts.

Round about the time of the optical disputes, Newton probably improved his earlier incomplete formulation of the binomial theorem and did other mathematical work, as well as spending some time on the physiology of vision, including a theory of binocular vision. However, he seems to have fallen round about 1680 into one of his frequent periods of distaste for science, for when Hooke, with whom he had become temporarily reconciled, wrote to him, at the end of 1679, asking him for a communication for the Royal Society, he says, in evading discussion: "But yet my affection to philosophy being worn out, so that I am almost as little concerned about it as one tradesman uses to be about another's trade or a countryman about learning, I must acknowledge myself averse from spending that time in writing about it which I think I can spend otherwise more to my content and the good of others." His expressed distaste for science may have had something to do with an event of considerable importance for his life which befell about this time: he met Charles Montague, afterwards Lord Halifax, who matriculated at Cambridge in 1679. The two men became intimate friends, and it was quite possibly Montague who first put into Newton's head that ambition to hold a public appointment which clearly was often with him from then on. Nevertheless, this interchange of letters with Hooke put in motion a train of events which led to the *Principia* being written.

Hooke in his letter had asked Newton what he thought of a supposition that he had put forward in 1674 in his *Attempt to prove the motion of the earth*, namely, that the motion of the planets could be explained by supposing a straight line motion which would continue in the absence of any force and a central force drawing the body aside, this force decreasing in some way with the distance. Newton ignored this question, but offered some considerations

* I am indebted to Mr. Orson Wood for calling my attention to these notes.

about the place where a falling body would, in consequence of the motion of the earth, strike the ground. He made a slip, which Hooke somewhat tactlessly corrected. Once more Newton was offended, and replied curtly, but Hooke, apparently unaware of the offence, wrote again and actually proposed quite definitely an attraction varying inversely as the square of the distance to account for the planetary motions. Newton took no notice, but it is more than likely that these letters revived his old Woolsthorpe interest in the problem of the celestial motions and set him thinking on the subject again.

Meanwhile, Hooke, Wren, and Halley are discussing the possibility of explaining the elliptical paths of the planets by the inverse square law, with Hooke, who is quite clearly convinced of the truth of the law, boasting that he can prove it, but unable to do so. Finally, Halley goes to Cambridge to see Newton about it, and finds that Newton has proved that a central inverse square law gives an elliptical orbit, but has lost the calculations. Soon after, Newton furnishes Halley with two proofs of the elliptical form of the orbit, and, his interest having apparently been renewed by working them out, writes down the elements of the mechanics of simple orbits in a little treatise, *De Motu Corporum*, which he afterwards sends to Halley.

De Motu Corporum embodied the contents of a course of nine lectures. In 1685 Newton began to write the *Principia* in extension of this work and established, which was essential for his elementary considerations of the moon's orbit, that with an inverse square law of force the mass of a uniform sphere can be considered as concentrated at the centre.* On April 28, 1686, the manuscript of the first book of this mighty work was presented to the Royal Society. The second book was finished by March, 1687, and the third and last by April of that year. Thanks entirely to Halley, who first of all realized the importance of Newton's work; secondly, agreed to "undertake the business of looking after it, and printing it at his own charge"; and lastly, when Hooke laid claim to the "invention" of the inverse square law, and Newton declared that he would suppress the third book, the crown of the work, persuaded him nevertheless to complete it—thanks entirely to Halley the book was published about mid-summer, 1687. It is one of the most prodigious feats of the human intellect through all time, and it was written in about eighteen months.

We know something of Newton's life at this time from Humphrey Newton, afterwards Dr. Humphrey Newton, who was no relation, but, as mentioned earlier, was constantly with him from 1685 to 1690. Unfortunately his reminiscences were not written down until after Newton's death in 1727; he says that if he had known that he would be asked for them he would "have taken a much stricter view of his life and action", that is, would have paid closer attention. He tells us that Newton very seldom went to bed "till *two* or *three* of the clock, sometimes not until *five* or *six*", and that he lay in bed only four or five hours. He often forgot to eat until reminded, and then ate a bit or two standing, drank very sparingly, and took no exercise, "thinking all hours lost that were not spent in his studies". When he was taking a short walk, "a turn or two" in his garden, he sometimes "has made a sudden stand, turn'd himself

* The proposition is, of course, true not only of a uniform sphere, but of any sphere for which the distribution of density is centrally symmetrical.

about, run up the stairs like another Archimedes, with an εὕρηκα, fall to write on his desk standing without giving himself the leisure to draw a chair to sit down on".

Humphrey Newton refers frequently to his chemical experiments called later by Stukeley "chymistry . . . that pyrotechnical amusement".* They must have been actively pursued during the five years 1685–1690: in the spring the fires in his furnaces scarcely ever went out. The transmuting of metals was his chief design, says Humphrey in one place, but elsewhere he hints at some mystical, some superhuman goal.

Evidently Newton forgot his irregular hours for, in his old age, he told Stukeley that "during his closest application, he never forgot going to bed about 12". This is but one of the many cases in which there are differing accounts, both apparently trustworthy, of events in Newton's life.

Shortly after the publication of the *Principia* a general upheaval was caused by the flight of James II from the throne in December, 1688. Owing to King James's interference there had already been trouble in the University, in which Newton had been honourably involved, and on the coming of William of Orange, Newton was returned as M.P. for Cambridge in the Convention Parliament, a dignity which he continued to hold for about a year. This seems worth mentioning, since attendance at Westminster may have given Newton a sight of the outside world and revived his anxiety to get a public place of some kind. At about this time Newton's mother died, which was a great shock to him. He seems to have been restless and anxiously seeking an administrative post: various friends tried first to have him made Provost of King's College, Cambridge, then to get him the office of Comptroller of the Mint, and after that to secure for him the Mastership of the Charterhouse, an appointment of which he did not think highly. The failure of all these advances made on his behalf was apparently a great disappointment. He had been relying upon the good offices of Montague, and we find him, at the beginning of 1692, writing: "Being convinced that Mr. Montague . . . is false to me, I have done with him . . .", the first sign of that period of melancholy and delusion, characterized by the conviction that everybody, including his intimate friends, was against him, which is often referred to as Newton's madness. This term gives, in my opinion, too grave a view of the disorder. It was towards the end of 1693 that these symptoms became really serious. At this time we find him writing to Pepys a letter about "the embroilment I am in and have neither ate nor slept well this twelve month", and saying that he must never see Pepys or the rest of his friends any more, a strange and causeless letter which alarmed the ex-President of the Royal Society. Within a few days of this letter he wrote to John Locke, accusing him in retrospect of trying to embroil him with women and by other means, and signing himself "Your most humble and unfortunate servant" in place of his customary "Your most humble servant". Locke returned a gentle, affectionate, and dignified answer, to which Newton replied that a distemper had put him out of order, that when he wrote before he had not slept "for five nights together not a wink", and that he did not know what

* "As to chymistry in general, we may very well presume Sir Isaac, from his long and constant application to that pyrotechnical amusement . . ."

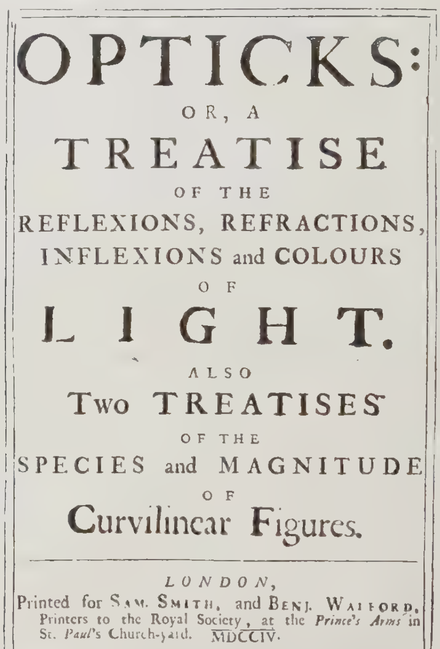
he had written. There is no doubt that Newton was, as he admits himself, ill, sleepless, and distempered, suffering from what is loosely called today a "nervous breakdown", but he was not, even temporarily, what can be called insane. The story that he was driven mad by the loss of certain manuscripts, which had been burnt by his dear little pet dog knocking over a candle during his absence, the story which tells of his affecting reproach, "'Oh Diamond! Diamond! thou little knowest the mischief done', without adding a single stripe", seems to be completely without foundation. As far as we can ascertain, Newton never had a pet dog; the fire, if in fact it took place, was some ten years previous to his derangement, and it was not of a nature to upset him.* The period of his illness and extreme strangeness was not very long; it would appear that by the end of the year 1693 his serious delusions of persecution had ceased.

It has often been stated, notably by Biot, that after this illness Newton lost his powers of grappling with scientific problems. To prove that this is quite wrong, we have three outstanding examples recorded, two mathematical and one experimental. In June, 1696, Jean Bernoulli published in the Leipsig Acts (*Acta Lipsiensia*) two challenge problems requiring a thorough knowledge of the calculus for their solution, one being the famous question of the brachistochrone, the curve connecting any two given points such that a particle sliding along it under gravity from one to the other shall take the shortest time in its descent. It was not until January 29, 1696/7, that Newton received a copy of the problems, and he gave a copy of the solutions to Charles Montague, the President of the Royal Society, on the next day. It was on receiving Newton's solutions, sent anonymously, that Bernoulli (recognizing the author from the style) exclaimed "*Tanquam ex ungue leonem*!"—that he knew him as the lion is known from his claw. A further challenge problem was set ten years later by Leibnitz "for the purpose of feeling the pulse of the English analysts". Newton received the problem on returning from the Mint and solved it before going to bed. So much for his failing mathematical powers. As for his acuteness as an experimental physicist, in the first edition of the *Principia* the outflow of water from a vessel through an orifice was incorrectly treated. In consequence of correspondence with Roger Cotes, in connection with the preparation of the second edition of the *Principia*, Newton reconsidered the matter, carried out further experiments, and discovered the phenomenon of the *vena contracta*, which he described in a letter to Cotes of March 24, 1710/11. The explanation has all his usual acumen.

In 1694 Newton was in close correspondence with Flamsteed about the latter's observations of the moon's motions, which Newton needed for work on extensions of his lunar theory. Both Newton and Flamsteed were lonely and difficult men, apt to see slights and to lack that delicacy of expression which characterizes our modern men of science. Their intercourse was marred by many misunderstandings before it was finally broken off. In March, 1696, Newton's whole mode of life suffered a profound change: he was made Warden, the second post, in the Mint, the first being that of Master of the Mint. To this latter office he succeeded in 1699.

* The rumour of a fire, once having been started, grew so mightily that Sturmius, the mathematician of Altdorf, reported that Newton's house and all his goods had been burnt.

Newton's work at the Mint was of considerable importance, and if he had never put forth any scientific work he would still have had something of the name that Samuel Pepys would have had if his diary had never been published—that of an actively efficient and faithful government servant at a time critical for the affairs dealt with by his particular office. A great scheme for the recoinage of all the money of England had been prepared by Montague, then Chancellor of the Exchequer, and Newton had to put the scheme through, which he did with great efficiency in circumstances of considerable difficulty. Newton was now a rich man, with influential friends, and he seems, with the help of his witty and beautiful niece, Catherine Barton, afterwards Mrs. Conduitt, to have entertained lavishly on occasion. Catherine Barton was on terms of some intimacy—concerning the exact nature of which there has been some indelicate debate—with Montague, Newton's old friend and promoter.


OPTICKS:

OR, A

TREATISE

OF THE

REFLEXIONS, REFRACTIONS, INFLEXIONS and COLOURS

OF

LIGHT.

ALSO

Two TREATISES

OF THE

SPECIES and MAGNITUDE

OF

Curvilinear Figures.

LONDON,

Printed for SAM. SMITH, and BENJ. WALFORD, Printers to the Royal Society, at the *Prince's Arms* in St. *Paul's* Church-yard. MDCCIV.

Figure 4. Title-page of Newton's *Opticks*.

After his appointment to the Mint, Newton spent little time and effort in science, though, even if he had lost his interest, he retained all his powers. Reference has already been made to his solutions of the challenge problems. The year after Hooke's death, which took place in 1703, he published his *Opticks*, most, if not all, of the work described in which had been completed long before. Remember that Newton was now sixty-one. He did not put his name on the title-page (see figure 4), the only explicit mark of authorship being the letters I. N. at the end of the *Advertisement*. I have never seen any conjecture as to a reason for this—possibly it was Newton's way of emphasizing that nobody but him could have possibly written it. In the fifteen queries which he gives at the end of the book, concerning debatable points, one deals with the theory of chemical

combination. This is one of Newton's very few published references to chemical problems. To the second edition of the *Opticks*, in 1717, which, with subsequent editions, bears his full name, he added fifteen further queries: they are of great importance for his views on light. The second edition of the *Principia*, which appeared in 1717, was prepared by Roger Cotes, but there is extant a lengthy correspondence which Newton carried on with Cotes about alterations to be made. A third edition of both the *Opticks* and the *Principia* appeared in Newton's lifetime.

In the last years of his life the position which Newton occupied was unique. His fame as a man of science—as a natural philosopher, to use the language of his time—was unrivalled throughout the learned world. Not only was he undisputed chief of mathematicians and experimenters in England: de l'Hôpital asked whether he slept and waked like other men. "I figure him to myself as of a celestial kind, wholly severed from mortality": Leibnitz, in 1701, said that "taking mathematicians from the beginning of the world to the time when Sir Isaac Newton lived, what he had done was much the better half". Halley's "Nec fas est proprius mortali attingere divos"—it is not permitted for a mortal to make a closer approach to the gods—is the first of a series of tributes conferring on him almost supernatural powers. Besides his scientific reputation, however, he held a great, lucrative, and honourable position: in 1705 Queen Anne had knighted him, which was an extraordinary honour for a man of learning in those days; society was open to him through his friendship with Montague and others; his niece was a toast of the town. From 1703 until his death, twenty-four years later, he was President of the Royal Society.* When he died in 1727 his body lay in state in the Jerusalem Chamber, and, as the funeral procession passed thence to Westminster Abbey, the pall was supported by the Lord High Chancellor, two dukes, and three earls. Such was the veneration in which Newton was publicly held that the place for his monument, allotted near the entry into the choir, was one that had "been often refused to the greatest of our nobility". James Thompson wrote an Ode on his death which contains many fine passages. Pope felt a reverence which he voiced in the famous couplet:—

> Nature and Nature's Laws lay hid in Night,
> GOD said, *Let Newton be!* and all was Light.

Voltaire, who always expressed the highest admiration for him, spent much effort in popularizing the Newtonian system in France. In the introductory poem to his book on Newton he even asks the "Confidens du Très-Haut, Substances éternelles" if they are not jealous of Newton—

> "Parlez, du grand Neuton n'étiez-vous point jaloux."

There has never been an Englishman of science or letters so acclaimed by his country and by the polite world in general.

It is meet that we consider briefly the peculiar glories of his achievement. In mathematics, certain of his discoveries, such as the bimonial theorem, are

* It is true that Sir Joseph Banks was president for a longer period, namely forty-two years, but he only retained his position by force of character and fighting. There was never anything but a unanimous wish that Newton should continue to hold the office.

known to bright schoolboys, but a difficulty in assessing his exact performance is that he avoided publication and, using his mathematics mostly as a tool for solving physical problems, seldom set down his methods systematically. There is no doubt that he knew the method of "fluxions" in 1666, and subsequently had no difficulty in solving problems involving the ordinary elements of the differential and integral calculus. It appears, from his results on the solid of least resistance, to which Dr. A. R. Forsyth has devoted special consideration, and from his solution of the problem of the brachistochrone, that he could solve problems which fall within the scope of the calculus of variations. He did fundamental work in the field of finite differences. With the geometry of conics he had a familiarity that few mathematicians to-day can boast of. An authority * has said that "Newton's tract is the first serious contribution to the Theory of Higher Plane Curves . . . It is a noteworthy sign of the versatility of his mathematical genius." Newton was one of the first mathematical geniuses of the world, but he seldom took the trouble to expound his advances. The whole sad and tedious dispute with Leibnitz about the invention of the calculus could not have arisen if he had published his findings in the ordinary way. There is now, I think, little doubt that Newton and Leibnitz were independent inventors. Newton could always find methods adequate for solving any problem which he had on hand. As an applied mathematician he probably had a surer instinct than anyone of his times: he seemed to smell the result. For formal advances in mathematics he displayed less interest.

In his work on optics Newton introduced a fundamentally new point of view, which men even as late and as intelligent as Goethe completely failed to understand. The scientific study of colour is extremely difficult, because the phenomena as they appear in nature are, in general, very complex, and our colour sense is very fallacious. Such subjective phenomena as contrast colours, simultaneous and successive, and appearances called forth by pressure of the eye: such colours as those produced by scattering in turbid media: the phenomena of additive and subtractive colour mixing—these are all fertile sources of confusion. Before Newton, nearly all talk on colour consisted of discussion of natural phenomena, uncritically considered, in terms of philosophical dialectic. Aristotle's views, in particular that colour is some kind of a mixture of light and darkness, or of white and black, appeared in various forms. Kepler begins: "Colour is light *in potentia*, light buried in transparent matter, if it is considered outside the line of vision", concentrating attention apparently, much as Aristotle and Goethe did, on a scattering medium. Grimaldi discusses whether light is a substance or an accident, but does say that colours are not something apart from light or really distinct from light. These examples may illustrate the way in which the nature of colour was being approached. Hooke, who clearly outlined a wave theory, but thought that light travelled faster in glass than in air, considered that all spectral colours were made up of a mixture of red and blue. Newton showed by his experiments with the prism that the degree or index of refraction could be taken as a quantitative index of colour—that the spectral colours were simple, in that no amount of refraction would change one once it had been isolated. He had great difficulty in making his point of view clear—

* Professor H. Hilton: *Isaac Newton: A Memorial Volume* (Bell, 1927).

that he was investigating how light behaved and not talking about what it was. His contemporaries mostly insisted on discussing what ought to be the behaviour of light on this or that theory. His own point of view Newton explained in the *Philosophical Transactions* in 1672, in reply to Father Pardies. "For the best and safest method of philosophising seems to be, first to enquire diligently into the properties of things, and of establishing those properties by experiments, and then to proceed more slowly to hypotheses for the explanation of them. For hypotheses should be subservient only in explaining the properties of things, but not assumed in determining them; unless so far as they may furnish experiments". The relation of theory and experiment could scarcely be more precisely expressed.

Newton rejected the wave theory put forward by Hooke because it would have, as a consequence, that light on passage through a small aperture should spread out. He did not recognize diffraction, which he had studied, as being just this spreading, but then neither did Huygens. He realized that the phenomena of thin films, on which Hooke had done fundamental work, never acknowledged properly by Newton, demanded periodic properties, and he inserted them into his corpuscular theory of light by his hypothesis of fits of easy reflection and refraction. This hypothesis assumed that the light particle went through periodically alternating states in which it was in turn easily reflected and easily transmitted by the surface of separation between two media. Newton suggested tentatively that waves generated at the surface by impact overtook the particles and governed their state. This theory would have been quite adequate for many purposes if the length of the fits did not have to depend upon the angle of incidence of the beam. Polarization compelled him to give "sides" to his particles. He missed seeing how transverse waves would include all the properties which experiment forced him to assume. So, however, did Huygens, whose wavelets had really no wave properties: they were only effective where they touched the envelope. He gave them this property to account for straight-line propagation, just as Newton assumed particles for the same purpose. Huygens' construction proved to be of the utmost value, but he was scarcely nearer to a true wave theory than was Newton: he gave his waves certain particle properties, just as Newton gave his particles certain wave properties. Even to-day we are not particularly happy on the wave and particle aspect of light.

Newton's work on light establishes him as a supreme experimenter. The same experimental skill is evident in, for example, the work on pendulums described in the *Principia*, the account of which must impress any experimental physicist by the care with which necessary corrections were carried out. Newton always knew exactly what precision a given experiment was capable of furnishing: "being not very sollicitous for an accurate calculus in an experiment that was not very accurate" is a typical remark. Lack of perfect sphericity in his globes, different stretching of the suspending thread when different weights were used, the effect of the sides of the vessel when the experiments were on the damping caused by different liquids—all such points come in for careful notice.

The heart of the *Principia* is the doctrine of universal gravitation. It is true, of course, that contemporaries of Newton had been feeling towards an explanation of Kepler's laws of planetary motion in terms of an inverse square

law of attraction, Hooke in particular having enunciated certain general principles for the solution of the problem. Galileo had done much towards Newton's first two laws of motion. Newton, however, started from the assumption that every *particle* of matter attracted every other *particle* with a force proportional to the product of the masses and inversely as the square of the distance: clarified the concept of mass, hitherto quite obscure, and showed that gravitational mass and inertial mass were the same: and then proved, by rigid methods, Kepler's laws of planetary motion, clinching the whole matter by his treatment of comets as visitors to our solar system. These services of his are generally realized. The additional matter, however, which the *Principia* contains is immense. Sir George Airy said: "If at this time we might presume to select the part of the *Principia* which probably astonished and delighted and satisfied its readers more than any other, we should fix, without hesitation, on the explanation of the precession of the equinoxes." Grant, the historian of Astronomy, selects Newton's researches on lunar theory as especially remarkable for their ingenuity and elegance. Professor Proudman tells us that it was the *Principia* that laid the foundation of all sound work on the subject of the tides. Newton was the first to solve the problem of the figure of the earth, and his result, reached by very simple considerations, is correct on his assumption, namely that of homogeneity. He was the first to form any idea of the true distances of the stars, which he did from the theory of gravitation. Laplace may be admitted as something of an authority on all that concerns celestial mechanics: it is recorded that, having in his hands the first edition of the *Principia*, he said to the Duke of Somerset: "This is the best book that was ever written", a statement quite in consonance with what he remarked on other occasions. Besides laying the foundation of celestial mechanics, the *Principia* also contains the foundations of much of mathematical physics, in particular of hydrodynamics. I need only mention the deduction of the velocity of sound, the effective definition of "Newtonian" viscosity in connection with his work on the Cartesian vortices, and his calculations about the vortices themselves.

The reflecting telescope and "Hadley's" quadrant, which he invented, are trifles to which a passing reference will suffice, trifles compared with Newton's other achievements, but enough to have made a lesser man remembered. Grant says of Newton: "His whole soul was wrapped up in the study of nature and her works". From a study of the *Principia* and the *Opticks* it might well seem to have been, but it certainly was not, in the sense that Grant meant. Newton had four main interests, and it is not easy to say that he spent more time and thought on one than on another. These interests were physical science, his achievements in which have won him universal glory; alchemy, in its mystical aspect; theology; and the work at the Mint and in the political world.

We have considered his work in the physical sciences. Alchemy occupied much of his time in the best years of his life. In 1675 Collins remarks that he was "intent upon chemical studies and practices": Humphrey Newton's reminiscences lay great stress upon his preoccupation with chemical experiments, and this during the period during which the *Principia* was written. His library contained a large proportion of books on Alchemy, and those which I have seen, such as Ashmole's *Theatrum Chemicum*, are heavily annotated by

him. His unpublished notes contain copious extracts from the alchemists of all ages. It would appear that he was greatly interested in the mystical side of alchemy, which would be in accordance with his character, for while he rigidly excluded any fanciful or mystical speculations from his work in the exact sciences, much of his theological work and his occupation with, for example, the works of Jacob Boehme and William Yworth, showed a strong element of mysticism in his character. Humphrey Newton seems to have suspected this, for he says of his chemical experiments: "What his aim might be I was not able to penetrate, but his pains, his diligence at these set times made me think he aimed at something beyond the reach of human art and industry."

On theological questions he was actively employed throughout his life. Conduitt records that Archbishop Tenison, when he offered him the Mastership of Trinity, which was in 1677, said: "You know more divinity than all of us put together." He was in active correspondence with Locke on theological questions round about 1690. Locke wrote to Lord King: "Mr. Newton is a very valuable man, not only for his wonderful skill in mathematics, but in divinity too, and his great knowledge in the Scriptures, wherein I know few his equals." In the catalogue of his library published by Colonel de Villamil there is a large number of theological works, including those of the early Church Fathers. His two works of theological interest published posthumously—*The Chronology of the Ancient Kingdoms amended* and *Observations upon the Prophecies of David and the Apocalypse of St. John*—bear witness to immense industry, knowledge and ingenuity. Nevertheless it is pretty certain that they would hardly be worth cataloguing by a bookseller of to-day if it were not for the author's name. A mass of unpublished theological writings also exists. They say Newton was a follower of the Arian heresy: we will leave such questions to the experts.

We have already referred to the part which Newton played in University politics and to his work on recoinage at the Mint. It may be said that this great scheme, which involved the setting up of mints in various parts of the country (in particular one at Chester, which was placed under Halley), was one of the first successful pieces of mass production. It was a great feat of technical organization. That Newton took this work very seriously seems to be shown by his letter to Flamsteed in 1669: "I do not love to be printed upon every occasion, much less to be dunned and teased by foreigners about mathematical things, or to be thought by our own people to be trifling away my time about them, when I should be about the King's business." In the first years of his appointment it undoubtedly fully occupied his attention. There is evidence that in 1710 he was very busy at the Mint, although in the last years of his life his duties there were light, and did not require him to attend more than one day a week, it not being the custom at that time to demand that one day's work should be spread over six days.

The strangest things have been said of his character. In his younger days there was much of the hermit about him. All through his life Newton showed an almost morbid dislike of any kind of controversy, a dislike which became, if anything, stronger as he grew older. In a letter to Halley he threatened to suppress the third book of the *Principia*, because "Philosophy is such an

impertinently litigious lady, that a man had as good be engaged in lawsuits as to have to do with her". He bitterly resented Flamsteed's criticisms. What was the cause of this hyper-sensitiveness is uncertain: it may be that, after, by the closest application and in the face of his keen self-criticism, he had won certain conclusions, he could not bear to have these conclusions put in doubt. He knew his own worth, and possibly felt that if he had satisfied himself, that was enough. It may be that his distrust and his aptness to see persecution had very deep roots in his character, in the dark soil far below the conscious. Be the cause what it may, of his suspiciousness and resentment of questioning there is no doubt. Locke, who seems to have been, as well as a great philosopher, a good-tempered and discerning man of the world, wrote that "he is a nice"—that is, difficult and over-precise—man to deal with, and a little too apt to raise in himself suspicions where there is no ground". Nevertheless, a legend grew up that he was particularly meek, and the poet Cowper says of him:

> "Patient of contradiction as a child,
> Affable, humble, diffident, and mild,
> Such was Sir Isaac".

And this "gentleness" is emphasized by Thompson.

Again, he was a shrewd man, who put through a very complicated and difficult piece of business in the recoinage, and, although clear of the taint of bribery, left a very large fortune. He was a very good business man. Macaulay, however, says of him, "with all his genius, he was as simple as a child". He had, of course, periods of intense concentration when he neglected his food and his sleep, but he was nothing less than the simple sage, gentle and unworldly, that popular legend has created.

The popular stories about him are mostly not only demonstrably untrue, but, what is more, quite out of character. We have already alluded to the "Diamond! Diamond!" fantasy. There is a story that he had two holes cut in his door, a large one to let through the cat and a small one to let through the kitten. Not only did he have no cat or kitten, but there is no reason to suppose that he liked draughts more than anybody else, and the man who built his own furnaces is scarcely likely to commit folly of this particular kind. There is a story that, carrying out certain simple calculations, he became so agitated because of his anxiety as to the result that he had to get a friend to finish them. Newton never showed any emotion of this kind; his only emotions, apparently, were those of anger and annoyance at anything that broke in upon his privacy.

Newton was a strange and remote character. He could mix in the world, but he was never of the world. Naturally something of a mystic and something of a recluse, at the same time he had, I think, an urge to show that he could carry out the duties of a public office as well as the next man. He could have performed with great efficiency any task where no emotional element was involved He aroused admiration, respect, reverence—but never love, hardly ever liking. His attitude to scientific controversy shows a shrinking from human contact. He wanted to find out, for his own satisfaction; he did not greatly mind telling people; but he would not argue with them. Wordsworth's

> "The marble index of a mind forever
> Voyaging through strange seas of Thought alone"

seems to me true, not in the sense that nobody had ventured into these strange seas before him, making uncertain casts and returning with doubtful tales, but in the sense that he sailed alone. He never discussed his work with anybody: he was willing to use the work of others, as of Flamsteed, but not to take them into his confidence. It is doubtful if he ever took anybody into his confidence.

Newton was not a man of constant aim, like Rembrandt, like Beethoven, like Faraday. He did not throughout his life think physical and astronomical science supremely worth while. He probably had greater powers of concentration than any other man: he himself said: "If I have done the public any service, it is due to nothing but industry and patient thought", and that if he differed from other men it was in his capacity to pay attention to a problem. Again, when someone asked him how he had arrived at his discoveries, he replied: "By always thinking unto them". At another time: "I keep the subject constantly before me, and wait till the first dawnings open little by little into the full and clear light." Everything goes to show that he could bring his whole mind to bear on a subject for hours on end, when he was interested in it. He did not, however, think physical science of all-important interest for more than a few years together. After the writing of the *Principia* he could be brought from time to time to turn his mind to it, but the zest seems to have gone. He could not bear that men of learning should think either that he could not solve a problem if he cared to try it, or that he had ever been indebted to a contemporary for one of his discoveries, but science for itself no longer occupied his thoughts with the possession that the beloved has over the young lover. He is one of the strangest and most baffling personalities among the very great. We greet his achievement with envy, his memory with reverence, but we do not understand him.

PHOTO-ELECTRIC ALLOYS OF ALKALI METALS

By ALFRED SOMMER, Dr. Phil.,
Cinema-Television, Ltd., London

MS. received 10 *February* 1941 ; *in revised form* 18 *December* 1942

ABSTRACT. The properties of the caesium-antimony alloy, first described by Görlich, have been investigated. It was found that this alloy has the stoichiometric formula $SbCs_3$, has the electric reistance of a semi-conductor and emits under favourable conditions about 1 photo-electron for 5 incident light-quanta at 4600 A. Alloys of similar structure were also investigated, and the effect of superficial oxidation was studied.

§ 1. INTRODUCTION

The photo-electric sensitivity of pure alkali metals is so small that very soon after the discovery of the external photo-electric effect, special sensitizing methods were developed to increase the sensitivity, these methods consisting of the introduction of other elements into the alkali layer. The first improvement consisted of sensitizing the alkali metal by a discharge

in hydrogen. Later, a special layer was developed, which is produced by distilling caesium on to oxidized silver with a subsequent heating process; this layer consists of a complex mixture of silver, silver oxide, caesium and caesium oxide. This layer, hereinafter termed the Ag—O—Cs layer, has been used for many years for most practical purposes because of its high sensitivity to radiation from normal light sources. The caesium can be replaced in this layer by other alkali metals, in which case the sensitivity is considerably reduced. Owing to the complex nature of the Ag—O—Cs layer, the spectral sensitivity and the quantum yield vary from sample to sample, and it is difficult to carry out exact experiments to elucidate the photo-electric mechanism of the layer. Figure 1 shows the spectral sensitivity curve of an Ag—O—Cs cell on an equal energy scale. (I am indebted to the National Physical Laboratory for carrying out the measurements on which figures 1 and 2 are based.) Two deductions can be made from this curve: firstly, the quantum yield at the maximum, at 8400 A.,

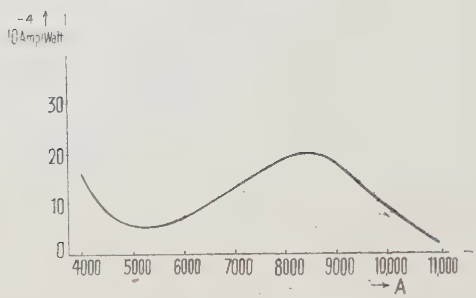


Figure 1.

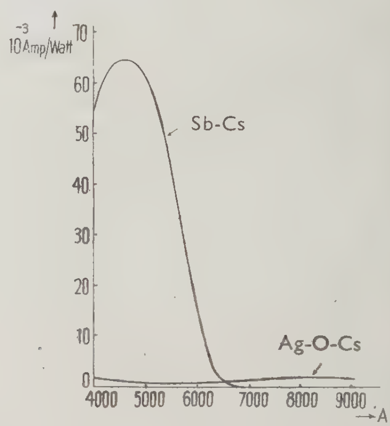


Figure 2.

is only of the order of 0·3 %. Secondly, the work function of the layer must be considerably below 1·5 volt because "infra-red" electrons of less than 1·5 volt velocity are able to leave the surface.

A new type of photo-electric layer containing an alkali metal was found by Görlich (1936) in the following way: he investigated alloys of caesium with other metals to obtain a layer which is uniform and sensitive from both sides (a useful property for transparent photo-electric layers in television tubes). Görlich found that alloys of Cs with metals like Bi and Sb fulfil this requirement. In a later publication Görlich (1937) emphasized that these *alloy layers* can only be used if they are very thin, a fact which he explains by the high resistance of the thicker layer, which prevents the supply of electrons to the surface layer. He obtained with a thick layer only about 10 % of the sensitivity of a very thin layer.

This effect and its explanation seemed to be improbable since a limitation of the photo-current caused by high resistance should make the sensitivity dependent on the illumination and on the voltage applied between cathode and anode. In our cells an effect of the high resistance was, in fact, observed, but

only wnen the illumination was very strong, corresponding to photo-currents of about $50\,\mu a$. In this case, deviations of the normal saturation curve occurred, but the effect was actually most apparent in very thin layers. This was to be expected, because the supply of electrons should be more affected by the longitudinal resistance of the layer (when deposited on glass) between the platinum contact and distant points of the layer than by the transverse resistance between the platinum and the surface. Hence a thicker layer. with lower longitudinal resistance, is likely to improve the supply of electrons. As to the sensitivity, we found that within the range of normal saturation the sensitivity of the layer at first rises with increasing thickness to become eventually independent of the thickness. This was to be expected because of the incomplete light absorption in very thin layers. Görlich's result is probably due to the fact that in his thicker layers the alloy did not have the optimum ratio of its components. The spectral sensitivity curve of a thick layer after superficial oxidation (see below) is drawn on an equal-energy scale in figure 2, and for comparison the curve for the Ag—O—Cs layer is redrawn on the same scale. A calculation shows that the quantum yield of the SbCs layer at the maximum, at 4600 A., has the surprisingly high value of 18 %, and, therefore, exceeds the quantum yield of the Ag—O—Cs layer by a factor of about 50, if one takes into consideration that a given amount of light energy at 4600 A. consists of fewer quanta than at 8400 A.

The SbCs layer appears to be more suitable for exact experiments than the Ag—O—Cs layer, because it consists of only two components and has a more uniform structure. This can be concluded from the fact that layers of very similar sensitivity can easily be reproduced, contrary to the results obtained with Ag—O—Cs layers. Therefore we investigated the composition and some of the properties of this alloy in the hope of determining the conditions for its photo-electric emission and for the extraordinarily high quantum yield.

The following problems were selected for investigation:

(1) How does the optical appearance (absorption and reflection of light) change during the formation of the sensitive alloy? The pronounced maximum of the sensitivity curve between 4000 and 5000 A. may be caused by preferential absorption of light in this range of wave-lengths.

(2) What is the ratio of caesium to antimony in the alloy? The alloys of alkali metals with metals of the fifth column of the periodic system have been studied in detail by Zintl (1931, 33). He found for Na and Sb intermetallic compounds of the stoichiometric formulae Na_3Sb_3, Na_3Sb_7 and Na_3Sb. Compounds of Cs have probably not been investigated before, but it seemed probable that the stoichiometric composition of the photo-electric alloy corresponds to one of these formulae.

(3) How does the electric resistance of Sb change during the formation of the alloy? This experiment should disclose whether the alloy has metallic character or not, and might also indicate, by sudden variations of the resistance, if alloys of different composition are formed.

(4) How does superficial oxidation affect the total and the spectral sensitivity? With the normal Ag—O—Cs layer we found in some cases a marked improvement of the sensitivity on employing superficial oxidation. If this oxidation is

not controlled very carefully, the sensitivity drops rapidly, after having passed through a maximum, because too many free Cs atoms, which are essential for the photo-electric effect, are oxidized. The initial effect of this formation of caesium oxide in the surface layer appears to be a lowering of the work function. The main evidence for this interpretation is the fact that the thermionic emission of the layer is greatly increased by the oxidation process. Therefore we can expect to obtain some information about the part played by the work function in the mechanism of the photo-electric effect of the SbCs layer by investigating the influence of oxygen on the electron emission.

(5) What are the properties of alloys in which Sb and Cs are replaced by metals of their respective groups? Görlich made alloys of Cs with other metals of the Sb group in 1936 and of Sb with other alkali metals in 1938, and has measured their spectral photo-electric sensitivity. In addition, we measured the stoichiometric composition and the resistance of some of these alloys in order to find a connection between these properties and the photo-electric sensitivity.

§ 2. EXPERIMENTAL METHODS.

The SbCs layers were produced on the inner wall of a spherical glass bulb of 80 mm. diameter (figure 3). The electrical connection to the layer consists of one or, for resistance measurements, two platinum tags P on the inner wall of the glass bulb. The antimony is evaporated from a filament F in the centre of the bulb to produce a layer of uniform thickness. This filament is also used

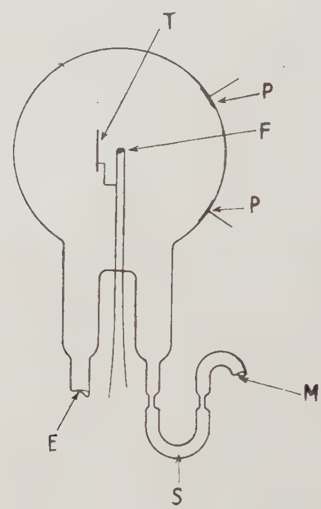


Figure 3.

as anode for photo-electric measurements. The bulb is sealed to the pump system at E. Caesium is introduced through a constriction from the side tube S. A window in the deposited antimony layer is obtained by fixing a small shield T near the filament.

To determine the ratio of Cs to Sb in the alloy, the following procedure is applied for quantitative experiments. A fragment of Sb of known weight is

fixed in the filament F and completely evaporated after the bulb has been evacuated and baked in the usual way. Thus not only is the total amount of Sb in the final layer known, but the thickness of the layer can also be calculated from the dimensions of the bulb. The thickness in most experiments was about 500 A. The amount of Cs in the alloy was determined in the following way: a glass capillary of bore about 1 mm., containing pure Cs, is weighed and put into an extension M of the side tube S before the whole system is evacuated. After the Sb layer has been deposited in the manner described above, the capillary is broken in two by a magnetically controlled iron ball and the whole of the Cs is distilled into S. The extension M is then sealed off and the two halves of the broken capillary are weighed so that the total amount (I) of Cs in the side tube S can be calculated as the difference of the weights of the capillary with and without Cs. Some Cs is then distilled from S on to the Sb layer in the bulb until the desired alloy is formed. Finally, the side tube S, containing the amount (II) of Cs that has not been used for the formation of the alloy, is sealed off and weighed. S is then opened, the Cs is removed and the side tube is weighed without Cs. The difference of the two weights gives the weight of Cs (II). The amount (III) of the Cs in the alloy layer is then calculated as the difference of the total amount (I) and the unused amount (II).

The longitudinal resistance of the layer can be measured between the two platinum contacts P, which are about 50 mm. from each other. The configuration of the deposit is such that the specific resistance cannot be directly deduced from the value of the resistance between the contacts, but the main point of interest is the ratio between the resistance of the pure Sb layer and the alloy layer. Furthermore, the specific resistance of the alloy layer can be calculated if one assumes that the known value of the specific resistance of Sb in bulk is the same as that of an evaporated layer. This assumption is most probably only correct to a limited degree, but we shall make use of it for the easier comparison of results obtained with different alloys.

The measurements of the total photo-electric emission were taken with an ordinary tungsten lamp. For measurements of spectral sensitivity, a red and a blue-green filter, for the wave-lengths below and above approximately 5700 A., were used in connection with a tungsten lamp. As a compensation for the lack of accuracy, this method had the advantage that a rough check on the change of spectral sensitivity was possible during the formation of the alloy and during the superficial oxidation.

§ 3. RESULTS

(*a*) *Observations during the formation of the alloy layer*

The distillation of the Cs on to the original Sb layer is accompanied by characteristic changes in the appearance of the layer. The first visible effect is that the Sb layer, which is practically opaque, becomes transparent and at the same time loses its metallic lustre. At this stage a very small photo-electric emission is first evident. If more Cs is introduced, the layer becomes more and more transparent, but the transparency is not uniform for all colours, since the layer now looks distinctly orange in transmitted light. The photo-electric emission does not rise appreciably during this colour change. After passing

through a maximum of transparency, the layer grows optically denser again and the photo-electric emission rises rapidly until it reaches a maximum. At this stage the layer shows a beautiful ruby semi-transparency. Further distillation of Cs does not affect the appearance of the layer, while the photo-electric emission starts dropping. The whole process of forming the alloy is accelerated by heating the glass bulb. During the whole process the ratio of sensitivity to "red" and to "blue-green" is practically constant, the sensitivity to "red" being almost negligible.

The transparency of the layer to red light, i.e. the absorption of the blue and green light, corresponds to the shape of the sensitivity curve (figure 2). Although the spectral absorption curve has not been measured accurately, the qualitative observation indicates that the absence of sensitivity to red light is due to the fact that red light is not absorbed.

(*b*) *Measurements of the Cs : Sb ratio*

The experiments were carried out in the manner described above. The following are the actual figures of three experiments :—

Weight (in mg.) of	Experiment No. 1	2	3
Sb evaporated	9·0	12·5	12·0
Glass capillary with Cs	202·5	216·5	234·5
,, ,, without Cs	158·0	160·0	173·5
Total amount of Cs (I)	44·5	56·5	61·0
Side tube + unused Cs	1400·5	1676·5	1347·5
,, ,, without Cs (after correction for weight of air)	1386·5	1658·5	1323·5
Unused Cs (II)	14·0	18·0	24·0
Used Cs (III) = (I) − (II)	30·5	38·5	37·0
Atomic ratio Cs : Sb $= \dfrac{\text{Cs (III)}}{\text{Sb evaporated}} \times \dfrac{\text{atomic weight Sb}}{\text{atomic weight Cs}}$	3·11 : 1	2·83 : 1	2·84 : 1

Assuming an accuracy of $\pm 10\%$, these results suggest that the photo-electric alloy corresponds to the formula $SbCs_3$, and is, in chemical nomenclature, caesium antimonide. Since this formula corresponds to the formula of Zintl with the highest ratio of alkali metal to antimony, it is quite possible that alloys of smaller caesium-to-antimony ratio are formed temporarily during the introduction of Cs into the layer. There was no point, however, in making a quantitative measurement during an intermediate stage, because the properties of the layer vary continuously during the formation process (see experiments with BiCs alloys, *infra*). On the other hand, an attempt was made to produce an alloy with a

higher Cs : Sb ratio than 3 : 1 in the following way. A cell, provided with a side tube containing excess Cs, was sealed off the pump system after the normal $SbCs_3$ layer had been formed. Then the Cs in the side tube was distilled on to the alloy layer, where it condensed as a metallic film, obviously without reacting with the alloy. Then the cell, together with the side tube, was heated to encourage a reaction, but without visible effect. Finally, the side tube was cooled while the cell was still at high temperature, with the result that the excess Cs condensed at once in the side tube. By weighing this amount of Cs in the usual way it was proved that no further Cs is absorbed by the alloy after $SbCs_3$ is formed.

(*c*) *Resistance measurements*

A typical result of a resistance measurement of the alloy layer (thickness of the original Sb layer about 300 A.) during the formation process is shown in the following table:—

Appearance of layer	Measured resistance (ohms)	Calculated * specific resistance (ohms/cm.³)
Metallic Sb layer	$7{\cdot}3 \times 10$	$4{\cdot}0 \times 10^{-5}$
Metallic lustre disappearing	$2{\cdot}5 \times 10^2$	$1{\cdot}4 \times 10^{-4}$
Layer of maximum transparency	$3{\cdot}0 \times 10^6$	1·6
Final $SbCs_3$ layer	$3{\cdot}0 \times 10^6$	1·6

* Assuming $4{\cdot}0 \times 10^{-5}$ ohm/cm.³ for the metallic Sb layer.

The table shows that the formation of the alloy $SbCs_3$ is accompanied by a rapid increase of resistance to about 10^5 times its initial value.

(*d*) *Influence of superficial oxidation*

The effect on the photo-electric emission of the $SbCs_3$ layer is similar to that achieved with Ag—O—Cs layers: the sensitivity rises to a maximum and then drops if the introduction of oxygen is continued. To obtain more detailed information about the influence of the oxidation on the spectral sensitivity, the oxidation process was carried out in steps, and the sensitivity to the three light sources, "white", "red" and "blue-green" (see above), was measured in each stage. The results of one experiment can be seen from the following table. The figures are given in % of maximum sensitivity to each light source.

Steps of oxidation	% maximum sensitivity White	Red	Blue-green
0	49	3·5	59
1	55	4·0	64
2	61	5·0	70
3	65	8·5	75
4	81	22·0	94
5	100	55·0	100
6	92	100·0	82
7	77	93·0	62
8	62	86·0	46
9	34	41·0	25
10	22	38·0	16

The first horizontal line of the table gives the values measured before the introduction of oxygen is started.

The table shows that

(1) The figures for "white" and for "blue-green" are of the same order during the whole sequence, and their maxima coincide. This was to be expected, because the absolute sensitivity to wave-lengths below 5700 A. represents the greatest part of the total sensitivity (see figure 2).

(2) The sensitivity to "red" rises by a factor of about 30 compared with the initial value, while the sensitivity to "blue-green" rises only by a factor of less than 2.

(3) The maximum for "blue-green" is passed before that for "red" is reached.

These results fit in very well with the assumption that the oxidation process lowers the work function. Lowering the work function must have a much greater effect on the "red" electrons, which have a very low velocity, than on "blue" electrons. With smaller work function, the sensitivity to red light will not only show a greater quantum yield, but will also be extended towards longer wave-lengths, since electrons with still smaller initial velocity will now be able to leave the surface. This explains the much greater improvement of sensitivity to "red" than to "blue-green" in the course of the oxidation. On the other hand, the $SbCs_3$ alloy, which is the actual photo-electric medium, will gradually be decomposed by oxidation (see below), and this unwanted effect will counteract the "positive" effect of reducing the work function. Therefore the sensitivity will drop as soon as the "negative" effect of the oxygen becomes greater than its "positive" effect. It is obvious that this turning point is reached earlier for "blue" electrons than for "red" electrons because the equilibrium of the two opposite effects will be reached earlier for the electrons of higher velocity, which are less affected by the "positive" effect of oxidation. This explains why the maxima for "red" and for "blue-green" do not coincide.

The amount of oxygen absorbed by the alloy layer is so small that an accurate measurement would be extremely difficult under the prevailing conditions. But the fact that the spectral-sensitivity curve after oxidation is still so fundamentally different from that of an Ag—O—Cs layer (see figures 1 and 2) indicates that only the very thin surface layer which determines the work function is affected by the oxidation, and not the actual photo-electric centres. The chemical reaction taking place after the introduction of oxygen can be expressed in the following way:

$$4SbCs_3 + 3O_2 \rightarrow 4Sb + 6Cs_2O.$$

Superficial oxidation has an effect not only on the photo-electric emission, but also on the electric resistance and the thermionic emission of the alloy layer.

With continued oxidation the electric resistance drops rapidly; this can be explained by the reformation of metallic Sb according to the above-mentioned scheme of reaction.

The effect on the thermionic emission corresponds to that in the case of the Ag—O—Cs layer (see above); while the pure alloy layer has an immeasurably small thermionic emission at room temperature, and also at higher temperatures.

an appreciable thermionic emission sets in with continued oxidation, which can be explained by reduced work function in the same way as in the case of the Ag—O—Cs layer.

(*e*) *Other alloys of the antimony-caesium type*

We tried all metals of the fourth and fifth columns of the periodic system in combination with caesium, and found that no other alloy has a photo-electric sensitivity as high as that of $SbCs_3$. The same applies for alloys of Sb with Rb and K. The alloy of As with Cs requires further investigation, because so far difficulties due to the low sublimation temperature of As have not been overcome.

We did not investigate each of these alloys to the same extent as the SbCs alloy, because of their lower sensitivity, but some results may be summarized as follows. (The experimental methods were the same as for the investigation of the SbCs alloy.)

(i) Bi + Cs. The formation of $BiCs_3$ was definitely confirmed. During the distillation of Cs on to the initial Bi layer, a transparent layer is obtained which changes rather suddenly into a final opaque layer. Since this stage is better defined than the corresponding stage during the formation of $SbCs_3$, a quantitative measurement was made, showing that this very transparent alloy layer has probably the stoichiometric formula BiCs and, therefore, corresponds to the compound Na_3Bi_3 found by Zintl and mentioned above.

Apart from the much lower photo-electric sensitivity (about 5% of the sensitivity of the $SbCs_3$ layer for the light of an incandescent lamp), the $BiCs_3$ layer differs from $SbCs_3$ by its more metallic behaviour: the layer has a slight metallic lustre, it is completely opaque at a thickness at which the $SbCs_3$ layer is transparent, and, a more conclusive fact, it has a much lower electric resistance. During the formation of the alloy, the specific resistance rises from $1{\cdot}19 \times 10^{-4}$ ohm/cm³ (pure Bi metal) to a maximum of 10^{-2} (transparent stage) and drops again to 2×10^{-4} when $BiCs_3$ has been formed.

(ii) Sb + Rb *and* Sb + K. The formation of $SbRb_3$ and SbK_3 was established. The photo-electric sensitivity of the two alloys is still lower than that of $BiCs_3$. In both cases an intermediate transparent layer is obtained during the formation of the alloy, corresponding to the transparent stages during the formation of $SbCs_3$ and $BiCs_3$. The specific resistance rises with increasing transparency from $0{\cdot}4 \times 10^{-4}$ (pure Sb metal) to 10^2 (for Rb) and to 10 ohms/cm³ (for K). The final layer of SbK_3 is slightly transparent to red light and has the specific resistance $1{\cdot}2 \times 10^{-3}$. $SbRb_3$ is more opaque, and its specific resistance is 10^{-4} ohm/cm.³. These results show that the alloys $SbRb_3$, SbK_3 and $BiCs_3$ are similar to each other, and different from $SbCs_3$, in having low resistance. It must be mentioned that the photo-electric sensitivity of the intermediate high-resistance stages of all investigated alloys is lower than that of the final layers. This is not surprising, because one cannot expect a high photo-electric emission if little or no light is absorbed, as is the case in these transparent high-resistance layers.

§4. SUMMARY OF RESULTS

(1) A photo-electric antimony-caesium alloy has been obtained which has a maximum sensitivity at 4600 A., corresponding to a quantum yield of 18 %, and a very small sensitivity above 6500 A.

(2) This alloy shows no metallic reflection and is very transparent to red light, the latter fact explaining the low sensitivity to this part of the spectrum.

(3) The composition of the alloy of highest sensitivity corresponds to the stoichiometric formula $SbCs_3$.

(4) The specific resistance of the alloy is of the order of one ohm per cm^3. This value is 10^5 times greater than that for metals and normal alloys. Hence the alloy belongs to the class of semi-conductors, in agreement with its non-metallic optical properties. As the term *alloy* usually implies metallic properties, it would probably be more correct to describe the substance $SbCs_3$ as an *inter-metallic compound*.

(5) Superficial oxidation has three effects on the alloy. Firstly, the photo-electric sensitivity is increased, particularly to red light. Secondly, the thermionic emission is increased. Thirdly, the specific resistance is reduced. The first two effects indicate a lowered work function; the effect on the specific resistance can be explained by the formation of metallic antimony.

(6) Alloys of antimony with rubidium and potassium and of bismuth and arsenic with caesium have been produced. They all show photo-electric sensitivity, but of a much lower order than the antimony-caesium alloy. Alloys of the formulae $SbRb_3$, SbK_3 and $BiCs_3$ have been obtained. As compared with $SbCs_3$ they have more metallic character: in particular their specific resistance is similar to that of metallic conductors.

ACKNOWLEDGEMENT

Acknowledgement is made to Messrs. Cinema-Television, Ltd., for permission to publish this information.

REFERENCES

Görlich, 1936. *Z. Phys.* **101,** 335; 1937. *Z. tech. Phys.* **18,** 460; 1938. *Phil. Mag.** **25,** 256.

Zintl, 1933. *Z. anorg. Chem.* **211,** 113; 1931. *Z. phys. Chem.** A, **154,** 1.

* Contain references to earlier work.

REVIEWS OF BOOKS

Physics and Philosophy, by Sir JAMES JEANS. Pp. 222. (Cambridge: the University Press, 1942.) 8*s*. 6*d*. net.

In this book Sir James Jeans returns once more to the exposition of modern physics and the discussion of its bearing on some of the traditional problems of philosophy. Everyone is familiar with his earlier efforts in this direction—*The Mysterious Universe*, *The New Background of Science*, and the Presidential Address to the British Association in 1934—but whereas these were all written within the space of four years, twice that interval has elapsed since the last of them appeared, and it might have been expected that the uncertainties of the earlier period would have given place to a more settled conviction. To some extent they have; a somewhat more definite viewpoint is now discernible, and it is less difficult to express the author's conception of the meaning of modern physics in a small compass. The book is nevertheless disappointing in that it indicates consolidation rather than progress. There is as little evidence as ever that Sir James sees anything in physics that has any actual bearing on philosophical problems.

The general summing-up of modern physics is as follows. There is something called " reality " with which we can never come into contact, the pattern as distinct from the essence of which is represented by mathematical expressions, including, for example, Heisenberg's formulation of the quantum theory. The particle and wave pictures of light and matter are partial representations of the truth embodied in the formulae. Neither representation is complete, but whereas the implications of the particle picture may or may not be true, those of the wave picture cannot be other than true. The waves do not represent electrons or photons, but our knowledge about electrons or photons. (This presumably means that one picture represents our knowledge of the other, for since electrons and photons are not mathematical formulae, and cannot be reality, which is inaccessible, there is nothing but the particles left for them to be.) The wave picture is completely deterministic and the particle picture indeterministic, but the indeterminism of the latter is not a property of " nature " (apparently a synonym for reality) but of our way of looking at nature, and the determinism of the former does not control events but our knowledge of events. Although it is stated that " the true object of scientific study can never be the realities of nature, but only our own observations of nature ", the precise relation between an indeterministic way of looking at nature and a deterministic knowledge of events remains somewhat obscure. Further, all our knowledge is only *probable* knowledge, and out of the possible alternatives we select the simplest as that most likely to be nearest the truth. The compatibility of this essential uncertainty with the complete determinism of our knowledge is another matter which would bear elucidation.

The outstanding stumbling-block of this scheme remains the difficulty of understanding why it is burdened with the essentially unknowable " reality ". If the word is too dignified and precious to be abandoned, why bestow it on something with which we can never come into contact and which can never enter into our consciousness? It seems an insult to such a noble conception. Why not start with " the true object of scientific study "—namely, our observations—and call them real? Experience seems a much more deserving recipient of the honour than nescience, and since we are inevitably eternally out of contact with the unknowable, it never will be missed if we cleanse our language of it.

This ghost of nothing is not merely an encumbrance to Sir James; it actually perverts his thinking by leading him in directions where he imagines it might make contact if it could. For example, when it is said that all our knowledge is only probable knowledge, some hidden inaccessible truth is implied, otherwise " probable " has no meaning. It is much simpler, therefore, to ignore the inaccessible and eliminate the adjective " probable ". But Sir James will not have this, and gives an example to show that the physicist's knowledge can only be probable. " He measures the wave-lengths of spectral lines in the light emitted by Sirius, and finds they are identical with those in the light emitted by hydrogen at a temperature of 10,000° C. He concludes without more ado that there

are atoms of hydrogen at 10,000° in Sirius. There is no proof of this and never can be, for we shall never be able to go to Sirius to find out." It is obviously implied here that if we could go to Sirius we could " find out ", and the knowledge would then cease to be probable and become certain. If that is so, then we can find out, by whatever process we would employ on Sirius, that a second sample of gas on the Earth which emits identical light is certainly hydrogen, and the proposition that we can have only probable knowledge is disproved. If it is not so, however, the illustration breaks down, for the knowledge obtained on Sirius is of the same character as that obtained on the Earth, and, therefore, affords no ground for calling the earthly knowledge probable. If Sir James would only try the effect of eliminating the superfluous from his considerations, he would be surprised at the simplification and illumination that would result.

The book has all the qualities that have made the author's earlier works so agreeable to read, and it deals effectively enough with some of the excesses of Eddington's philosophy, though there seems to be room for a deeper appreciation of the considerations which have led to those excesses. In case Mr. A. P. Herbert misses the book, we would enter a mild protest against the word " lecturee " on page 76. H. D.

Molecular films, the cyclotron and the new biology, by HUGH S. TAYLOR, ERNEST O. LAWRENCE and IRVING LANGMUIR. Pp. viii + 111. (New Brunswick, N.J.: Rutgers University Press, 1942.) $1.25.

There can be few branches of applied physics which show more promise and liveliness at the present time than that lying on the borderland between physics and biology. Just as other applied physics passes insensibly into engineering, so does biophysics into physiology; but there seems to be developing a branch of applied knowledge distinct from physiology, though adjacent to it, and the essays under review give a vivid impression of this new field in its cultural, philosophical and most brutally experimental aspects.

Dr. Taylor's essay provides an historical background covering a very wide range of physical and chemical discoveries which led to the development of biophysics and its utilization of new techniques of amplification, electronics, recording by photocells, oscillographs, thermopiles and sound-meters. His essay ends with an inspiring call for an integration of spiritual forces and a plea for the mutual co-operation and assistance between individual sciences; a call for an increasing breadth of culture and of education among scientists themselves, and for a dedication of the noblest minds to the forward march of knowledge with full awareness of the social consequences of that knowledge and a franker recognition of the other factors contributing to wisdom.

From such considerations we turn rather abruptly to the technical details of Langmuir's essay on molecular films in chemistry and biology. Short, simple, definite sentences ; clear-cut simplified conceptions of molecular problems ; diagrams of simple troughs and spreading films of proteins. The fundamental information thus furnished, or perhaps about to be furnished, is astounding. What happens if you put a drop of oil in water ? What happens if you touch a minute fragment of crystalline pepsin on to that ? Often the writer breaks off with " I do not know the significance of these things " ; " I would like you to believe in the reality of these concepts " ; " I like to think of the concreteness of them ". " The rapidity with which you can study phenomena is extremely great because of the simplicity of the method ". A masterly essay.

From such superb simplicity we proceed to some of the most complex and costly investigations ever undertaken, the cyclotron, growing ever more intricate and more powerful. Lawrence sketches very briefly and with enthusiasm the achievements and prospects of artificial radioactivity, the neutron and its biological action, including the clinical effects of beams of neutrons. In his judgement, neutron therapy will eventually take an important part along with surgery, X rays and radium in the treatment of cancer, Much of the essay is concerned with " synthetic radioactive tracer atoms ", as, for example. in the study of the thyroid by radioactive iodine, or the fascinating story of the use of element 85, which was predicted to be like iodine and found to be taken up by the thyroid much as is iodine. Other sections deal with calcium and strontium metabolism ; " radio-autography ", in which the distribution of, for example, radioactive zinc is shown very clearly by slicing the fruit and placing the slices against a photographic plate ; the treat-

ment of leukemia by radiophosphorus, osteogenic sarcoma by radiostrontium, and the study of photosynthesis by radioactive carbon. Description and a number of illustrations of the giant cyclotron and its workings are given. Finally, commentaries on the present situation by Chambers and Dunning are reported.

It is a great pity that Lawrence's article is marred by absurd mistakes in the illustrations. For example, figure 11 shows a middle-aged man who has been subjected to a beam of neutrons for a " carcinoma involving extensively the jaw bone ", and certainly as far as can be seen the reaction is very pronounced. The legend asserts that " The next figure shows the patient several months later. The skin has healed, and the tumour has evidently disappeared, being replaced with scar tissue ". The reader, in anticipation, turns over to find a picture of a charming little Chinese girl, probably about twelve years of age. Clearly, in spite of the protestations, neutrons *can* achieve miracles ! Again, figure 19 purports to show the comparison of the accumulation of radiophosphorus and radiostrontium in the femurs of two rabbits. " The radioautograph on the left indicates that radiostrontium was predominantly accumulated in the marrow ". In fact, the one on the left is labelled Sr^{89} and shows precisely the reverse. Such mistakes are perhaps trifling, but they are unsettling to those anxious to make a just estimate of the care and reliability of the whole of this very difficult but important work.

Finally, this volume, published by the Rutgers University Press (New Jersey), in honour of its one hundred and seventy-fifth anniversary celebration, is beautifully produced and printed, and gives a highly stimulating account of the way in which the applications of physics in biology and medicine proceed apace across the Atlantic. W. V. M.

Life of Sir J. J. Thomson. By Lord RAYLEIGH. Pp. x + 300. (Cambridge: the University Press, 1942.) 18*s*.

The *Life of Sir J. J. Thomson*, by Lord Rayleigh, has now appeared from the Cambridge University Press—a handsome volume bearing few, if any, physical signs of war-time restrictions. One opens it with eagerness, not unmixed with some little misgivings. To produce a biography of Sir Joseph Thomson, O.M., sometime Cavendish Professor of Physics and later Master of Trinity College, Cambridge, would be comparatively easy—any competent official biographer could do it ; to write a life of J. J. which would satisfy the generations of students who have revered and loved him, and who owe to him some of the happiest days of their lives, is quite another matter. Not half the greatness which we knew in him was revealed in his science. Could any pen portray to future generations the authentic likeness of J. J. as we knew him, set against the essential background of the research school which was his own creation, and perhaps his greatest contribution to science ? Could any writer make clear to a generation for which the electron has become almost a commonplace domestic appliance the genius which went to its discovery, or recapture, after a lapse of forty-five years, the intellectual excitement of the days described by the Cavendish lyricist :

" All preconceived notions he sets at defiance
By means of some neat and ingenious appliance,
By which he discovers a new law of science
Which no one had ever suspected before " ?

These are questions to which we look to this volume for an answer. Its author is peculiarly well equipped for the task. He is a Trinity man (and Trinity meant much to J. J.); he has already proved his quality as a biographer in his distinguished *Life of Lord Rayleigh*, and, above all, he was attached to the Cavendish Laboratory, as a somewhat precocious undergraduate, at the time of the discovery of the electron, and took an active part in the hunt which followed. To this last fact we owe the following characteristic vignette :

" In the summer of 1897 J. J. was bubbling over with enthusiasm over his work on cathode rays. The first I heard of it was from himself. I was at the time only an undergraduate, but he knew, I think, from the questions I had asked him after his lecture that I was as eager to hear as he could be to talk, and chancing to meet on King's Parade he began to unfold to me what he had been doing—telling me that the cathode rays had now ' turned out ' to be particles, and particles quite different from atoms. My rooms

were at that time in Whewell's Court, but I did not want to interrupt the tale he was unfolding by stopping there, and walked on with him past St. John's and the Round Church to the other entrance of Whewell's Court in Sidney Street, where he left me, after standing talking for a few minutes."

There is much of the essential J. J. in this little incident.

It is by a series of such sketches, rather than by any formal analysis, that the author builds up, step by step, the many varied aspects of the personality of his subject. One would be tempted, did space permit, to quote at large. Two or three examples must suffice. The first will recall happy memories to all old Cavendish students.

"The tea hour was in many ways the best time in the laboratory day. The tea itself had no special quality ; the biscuits were unattractive in the extreme, and very dull ; but the conversation sparkled and scintillated, and as a social function tea was an outstanding success. There seemed to be no subject in which J. J. was not interested and well informed : current politics, current fiction, drama, University sports, all these came under review. . . . J. J. had something to say on nearly any subject that might turn up. He was a good raconteur, but also a good listener, and knew how to draw out even shy members of the company."

Or, again, another characteristic picture :

"His mind was working incessantly. If an idea struck him in the laboratory, he would bend forward a good deal, rub his hands together vigorously and dart across the room as far as the size of it would allow. Perhaps it was this impact of a new idea which would suddenly make him change his motion from a walk to a jump when he was going along King's Parade with his hands in the pockets of the veteran overcoat, sometimes chuckling to himself."

And lastly, a shrewd comment on Thomson's scientific method.

"The story is often told that Newton laid aside the calculation in which he connected the moon's motion with terrestrial gravity because there was a discrepancy of the order of 10 per cent due to a wrong value of the earth's radius. This story is not universally received . . . . but whatever Newton may have done, it is very certain that Thomson would not have abandoned his thesis even temporarily in like circumstances."

The method, though perhaps a little unusual for an official biography, succeeds to admiration ; one doubts whether with a character so rich and varied as that of J. J. any other could have produced an equal effect. The dangers of mere anecdotage are carefully avoided ; every reminiscence makes its point, and adds its quota to the full portrait ; and the background of the Cavendish research school—the first in history—is sketched in with a sure touch. Perhaps a reviewer who enjoyed for nearly three lustra the high privilege of working under J. J. is not in the best position to say how the resulting picture will appear to readers to whom he is only a legend. I can only say that to me it seems both vital and authentic, and express my gratitude to the author for embodying it in lasting form.

The biography of a scientist must be a scientific book. It must attempt to expound the scientific discoveries which he made, their relation to the science of the time, and their importance for the science of the future. This is by no means easy. If the account is too technical it will be incomprehensible to any except the expert, who would, presumably, know all about it already ; if too popular it ceases to convey any real impression.

Lord Rayleigh has tried to steer a course between the two extremes. To quote from the Preface : "No formulae have been used, in view of the protest of a well-known literary man, that he could not even skip a formula ! " On the other hand, no concessions have been made to weakness, where they might detract from the clarity of the argument. Physicists might study with profit this demonstration of how very far it is possible to go in physical reasoning without the use of mathematical paraphernalia. In view of the controversies as to priorities, faint echoes of which still linger in the air, this authoritative and impartial account, by one who was in an excellent position to know the facts, of the work done during the great days at the Cavendish, is a real contribution to the history of science.

Great biographies are rare, and only the test of time can show whether this *Life of Sir J. J. Thomson* will ultimately rank among them. Lord Rayleigh can, at any rate, be congratulated on having produced a record which will not only find its place on the physicist's book-shelf, but will also be very often in his hand.

J. A. CROWTHER.

Electrodynamics, by LEIGH PAGE and NORMAN ILSLEY ADAMS. Pp. xii+506. (London: Chapman and Hall, 1941.) 32*s*.

Unless a treatise on electromagnetism presents a new theory which, like O'Rahilly's, purports to revolutionize the subject, it is bound to be in the nature of a restatement, especially if, as in the work under review, the more or less recent developments introduced by the quantum theory are not included. But this treatise is a clear, bright, invigorating restatement, which cannot fail to appeal to students of the subject, and may rightly be numbered among their favourite text-books.

The work starts with a useful introduction to vectors and their application to the study of potential, terminated by a short section on tensors. An exposition of the principle of relativity follows, which is not too easy reading, except possibly for those well steeped in the subject.

A so-called emission theory of electromagnetism is presented in the third chapter: each line of force is considered as the locus of points moving in straight lines with the velocity of light. The authors themselves indicate, however, that this theory starts with the elementary electric field as the fundamental entity in its description of electromagnetic phenomena: it is therefore not a ballistic theory, and is an emission theory in a restricted sense only. From this hypothesis are deduced expressions for the scalar and vector potentials, as well as for Maxwell's equations. It is noteworthy that the relativity formulae given in the preceding chapter are not used for this derivation of Maxwell's equations. The equation for the force on a charge in motion in electric and magnetic fields is derived in a similar manner, and generalized to include magnetic as well as electric charges.

From then on we enter the realm of applications, and interest is not only sustained but grows throughout the rest of the book, as explanations of well-known phenomena (Thomson, Hall, Ettinghausen, Righi-Leduc, Nernst) are deduced, and the mathematical treatment unfolds before the reader the general solutions of familiar problems.

The electromagnetic wave equations are obtained from Maxwell's equations in the ordinary way, and there follows a complete and very clear exposition of the propagation of waves, the discussion on wave guides being particularly instructive, as is also that on radiation from antennas in the next chapter.

Then comes an illuminating discussion of light waves, in which the phenomena of transparency, scattering, reflection from dielectrics and metals, transmission through crystals, and optical activity are discussed; the treatment of group velocity is worthy of special mention.

What has been said of the exposition of the principle of relativity applies to the treatment of four-dimensional vectors. The last chapter, on general dynamical methods, including a derivation of the canonical equations, is of more special interest, followed as it is by applications to practical problems like the magnetron and cosmic-ray trajectories, which conclude a treatise which the publishers rightly claim will become indispensable to those who are fortunate enough to possess it.

The greatest praise is also due to the publishers for the printing and presentation, and to all concerned for the scarcity of misprints. P. V.

Electrical Counting, by W. B. LEWIS. Pp. viii+144. (Cambridge: the University Press, 1942.) 10*s*. 6*d*. net.

While everyone appreciates in a general way the power and flexibility of the thermionic valve as a physical instrument, the fullest use of this instrument will not be made unless those who have problems to solve are kept aware of what it can do; a special welcome is therefore due to any book which authoritatively describes the use of valve circuits for other purposes than those of electrical communication.

The skeleton of Dr. Lewis's book is the technique of counting ionizing particles, but on this framework he hangs a substantial body of information on the properties of valves and the circuits in which they may be used to amplify and count voltage impulses as small as 10^{-5} v., recognizing as separate events two impulses whose instants of commencement may differ by perhaps 10^{-4} sec., but whose total duration may be many milliseconds, and discriminating, if necessary, between impulses of different magnitudes. The

explanations of the circuits are concise and clear ; the comments on their advantages and limitations, and the practical hints as to their use, are such as can only be given by one who has had much experience in work of this kind. The chapters which deal with the principles of valve amplification (including the use of feed-back) are very well done ; added to Dr. Lewis's habit of giving an elementary explanation of every important circuit used, and to his liberal use of annotated circuit diagrams, they make the book a very useful auxiliary manual on low-power valve circuits at frequencies up to about 10^5.

Those who are interested in nuclear physics will find much of value in the sections on ionization chambers for counters, on Geiger-Müller counters, on range and energy determinations and on the statistics of random distributions. This last subject is one where, if the reader is to follow the exposition without continually checking the statements he meets, the author must not only be right but must inspire confidence in his correctness ; it is, therefore, unfortunate that (as on page 116) the term " probable number " is used without explanation in the sense of *average* number. The reader will be much puzzled if he should think, as the reviewer at first did, that " most probable number " was meant.

The descriptions of the various available electro-mechanical devices for displaying the result of a " count " are, though not detailed, quite adequate for an assessment of their merits.

The chapter on the limitations of amplifiers presents an admittedly unsatisfactory picture of the physics of valve and circuit noises ; here, in particular, it is a pity that Dr. Lewis could not include work published since the war began. It is no disparagement of a most valuable book to express the hope that the author may, in a second edition, be able to bring it well up to date ; meanwhile we must be grateful to him for finding time to share with us his wide and deep knowledge of electrical counters. P. B. M.

RECENT REPORTS AND CATALOGUES

Recent Books on Wireless and Physics. Pp. 2. Chapman & Hall, Ltd., 11 Henrietta Street, London, W.C. 2.

Overhead Lines Loading Tables. Pp. 10 (12 in. × 11 in.). The Copper Development Association, Grand Buildings, Trafalgar Square, London, S.W. 1. *Gratis.*

Notes on the Value of Microcopying to Research Organisations. Pp. 3. Ilford, Ltd., Ilford, Essex. *Gratis.*